Report No. P67-15
HAC Ref. No. A8468

ADVANCED DEEP SPACE COMMUNICATION SYSTEMS STUDY

Contract No. NAS 12-81

Final Report

January 1967

Prepared by
Research and Development Division
AEROSPACE GROUP
Hughes Aircraft Company • Culver City, California
for
NASA Electronics Research Center
Cambridge, Massachusetts

PARTIAL LIST OF CONTRIBUTORS TO THIS REPORT

E. C. Park, Senior Staff Engineer, Research and Development Division, Program Manager

L. L. Bailin, (on leave of absence)

J. F. Cashen, Member of the Technical Staff, Space Sciences Division

D. C. Forster, Manager, Electron Device Department, Research Laboratories

R. R. Gold, Manager, Aerospace Physics Research Staff, Space Systems Division

S. D. Hamren, Manager, Antenna Department, Research and Development Division

G. Hrycenko, Member of the Technical Staff, Research and Development Division

W.H. Kummer, Senior Scientist, Systems Laboratory, Research and Development Division

H. R. Senf, Head, Applied Quantum Electronics Section, Research Laboratories

A. F. Seaton, Senior Technical Staff Assistant, Research and Development Division

S. S. Shapiro, Member of the Technical Staff, Research and Development Division

L. S. Stokes, Senior Scientist, Systems Laboratory, Research and Development Division

A. T. Villeneuve, Senior Staff Engineer, Research and Development Division

CONTENTS

ILLUSTRATIONS

TABLES

0.0 PROGRAM SUMMARY

The primary objective of the Advanced Deep Space Communication Systems Study has been to define those areas of technology where research effort can best be expended in order to meet deep space communication needs in the period 1970 to 1980. Accomplishment of the primary program objective has required the establishment, first, of fundamental system limitations to the extent that they are understood as well as the practical restrictions imposed by current technological limitations. Based on this information, potential system performance has been estimated as a function of frequency to serve as a guide to the selection of optimum frequencies and to provide an indication of the improvement required at each frequency. Finally, the required performance improvement has been apportioned among the various system parameters in accordance with estimates of present and projected technological progress in order to identify the parameters and corresponding technologies for which the greatest gain in system performance can be expected for research effort invested.

This report discusses first, in Section 1, the communication system requirements for deep-space missions with reference to transmission range, data rate, and angular coverage. The high data rates to which this study is directed are appropriate to photographic and other mapping missions.

Basic limitations to system performance are reviewed in Section 2. These include minimum signal requirements, noise sources and their spectral characteristics, and atmospheric attenuation and distortion effects. Practical limitations on performance due to the present and projected state of technology are discussed later with reference to radio frequency and optical components in Section 4 and 5, respectively.

Based on the various performance limitations expected communication link performance is then analyzed in Section 3 as a function of frequency to determine suitable frequencies and system configurations. Several candidate systems are described.

As mentioned above, Sections 4 and 5 are concerned with a review of the present state of the art in radio frequency and optical technology as well as with a discussion of potential developments and promising areas for research.

Plasma propagation as related to communications during atmospheric entry is considered in Section 6. Communications "blackout" is discussed as a function of elapsed time and altitude bands for several frequencies of interest.

Section 7 summarizes the many areas in which research effort could usefully be applied to improve space communications capabilities. Areas of particular pertinence or promise are called out and further discussed.

0.1 SYSTEM PERFORMANCE LIMITATIONS

Signal coding techniques are available which permit a close approach to the theoretical data rate limit. Similarly, receiver performance in terms of noise and detection efficiency is approaching reasonable or ultimate limits within the spectral region of interest except at millimeter and submillimeter wavelengths. Again, efficient narrow-band radiative sources capable of suitable CW power levels are available in both radio and optical regions of the spectrum, except at submillimeter, near-infrared and visible wavelengths. (Two years ago the entire optical region would have been excluded.) The significant restrictions on system performance are therefore due to external noise sources and atmospheric effects which identify certain favorable regions of the spectrum, and to engineering and technological limitations on aperture gain which determine the achievable levels of performance within these favorable regions.

Galactic noise and atmospheric background tend to define a low-noise region between 1 and 10 GHz under fair weather conditions, which is narrowed to about 1 to 5 GHz during rain. Above 50 GHz atmospheric absorption becomes prohibitive out to a wavelength of about 12 microns except for a partial window at 94 GHz; and cloud

cover extends the blackout through the optical region. Thus for a ground-based receiver good performance is limited to the 1 to 10 GHz region and optical wavelengths; and with rain and clouds the useful frequency band is further restricted to 1 to 5 GHz.

Distortions of the incident wavefront due to atmospheric turbulences set limits on the effective dimensions, and hence gain, of single-element receiver apertures. In the microwave region, gains are limited to the order of 80 db, and in the optical region to the order of 100 db. Achievable fabrication tolerances set gain limits in a similar manner on both receiver and transmitter apertures. A given ratio of rms deviation in effective pathlength to aperture diameter results in a gain limit constant with frequency. However, in practice smaller tolerance ratios are achievable for smaller diameters so that as aperture dimensions decrease with increasing frequency, an improvement in gain can result. Gain is finally limited at the high-frequency end of the spectrum by the need for a beamwidth comparable with expected pointing errors.

0.2 CANDIDATE SYSTEMS

The basic limitations on system performance define two spectral regions of interest: one in the microwave region in the vicinity of 3 GHz, and the other in the optical region. However, since poor weather conditions can effectively black out the entire optical region, the requirement for an essentially continuous optical link would necessitate an increase in the number of ground stations at diverse locations to assure fair-weather conditions for at least one, or an earth-orbiting satellite relay station.

For a ground-based optical receiver atmospheric distortions of the wavefront restrict single-aperture diameters in an approximate inverse relationship to frequency. Thus multi-element apertures are required to provide equivalent performance at higher frequencies, and receiver simplicity tends to favor the longer optical wavelengths. The unique availability of an efficient transmitting source at 10.6 microns provides

an overriding argument which directs interest to the 10-micron region for an optical communication link employing coherent reception.

For noncoherent reception, receiver sensitivities as limited by thermal or quantum noise are sufficiently below coherent levels to require excessively large aperture areas. Although these "photon buckets" need not be of optical quality, practical limitations on detector size would require fabrication tolerances comparable to those required for microwave apertures with the additional requirement of a specular surface. Therefore, unless an efficient laser becomes available in the visible region, noncoherent reception is not recommended for wide-band communication links.

For a satellite receiver atmospheric restrictions do not apply. However, the large aperture dimensions appropriate to microwave frequencies, with no appreciable compensating advantages in terms of reduced atmospheric losses and background noise, eliminate a microwave spacecraft-to-satellite link from further consideration. At the same time the millimeter-and submillimeter-wave region offers little advantage over the optical region with respect to the state of the art of radiative power sources and detectors, and the larger apertures required put this region at a comparative disadvantage. The optical region is therefore favored. Gain limitations due to pointing accuracies and the present availability of an efficient transmitting source at 10.6 microns once more direct attention to the 10-micron region.

Three basic system configurations are therefore considered worthy of further investigation. These are:

- a direct microwave, spacecraft-to-earth communication link in the region of 1 to 5 GHz;
- a direct optical link at 10.6 microns employing additional ground stations to assure the necessary weather diversity;

- a satellite relay configuration (one or at most two required) utilizing a 10.6 micron optical link from spacecraft to satellite with a noncritical microwave link for the short range from satellite to earth.

Further analyses are required to determine the relative costs appropriate to these configurations.

0.3 RECOMMENDATIONS FOR RESEARCH PROGRAMS

The present study has narrowed the choice for a deep space communication link to three candidate systems. However, a quantitative evaluation of these three configurations is not within the scope of this program. Considerable further study and detailing of system components are required before a valid comparison can be made. In Section 7.6 several study programs are recommended, having as their goal the final selection of configuration and a conceptual design of the communication system. Some preliminary research and development of the critical components in each system is necessary to provide realistic cost estimates, and is desirable to maintain technological progress through the selection period.

The primary program recommended (Section 7.6.1) is essentially an extension of the present study. A comparison of system parameter values as influenced by fundamental and practical restrictions on performance has reduced the number of promising configurations to three. A more detailed analysis of each system and its critical components is now required to evaluate the candidates quantitatively. The evaluation process can be greatly facilitated by the use of a computer program which can at each stage of the study determine the optimum division of burden among the system parameters. Thus, simultaneously, attention will be directed to the significant areas for investigation, and a running account of estimated costs, including necessary R and D efforts, will be kept.

In order to provide more valid and reliable data for the systems evaluation study, programs should be initiated concurrently to consider

the conceptual design and methods of reducing costs of critical system components. The critical component studies are summarized in Sections 7.6.2 to 7.6.4 with reference to the three system configurations and comprise:

- a study of optimum methods and techniques for the implementation of large ground antenna arrays.
- an evaluation in terms of performance and burden of spacecraft antenna configurations, including simple antennas, extensible antennas, and multi-element phased arrays.
- a study of coherent optical receiver techniques including heterodyne frequency and phase control and electro-optical pointing and tracking.
- an investigation of basic laser mechanisms and frequency control of the carbon dioxide laser.
- a study of acquisition, tracking, and pointing requirements on board the spacecraft including the optimum relationship between transmitter beamwidth and pointing accuracy.
- a comprehensive investigation of atmospheric distortion effects on coherent reception with reference to site and weather.

Concurrent with these studies which are concerned with the definition of an optimal system design, it might be advantageous to pursue the development of high-gain extensible spacecraft antennas. Such an antenna might give a 10-20 db increase in performance pending development of a large microwave ground receiver array or an optical communication link.

1.0 SYSTEM REQUIREMENTS

Deep space missions for the forseeable future will be restricted to exploration of the sun, the planets and their natural satellites, and to the space environment within the dimensions of the solar system[1,2]. For the time period (1970 to 1980) and the data transmission rates toward which this study is directed (10^6 to 10^8 bits per second), missions of interest will be restricted to orbiter or lander explorations of the nearer planets. These missions define the basic system requirements with reference to transmission range, data rate, and angular coverage. The communication path of primary interest is that from the space probe to Earth both because it represents the essential direction of data flow and because vehicle restrictions on weight and available prime power cause this link to be the most critical.

1.1 TRANSMISSION RANGE

Communication range for a given space mission depends on the launch date and the injection energy expended as well as the objective. Figures 1-1 through 1-4[3] show communication distance vs. launch date with C_3, the injection energy of the escape hyperbola, as a parameter for Mercury, Venus, Mars, and Jupiter missions. Of most immediate interest are missions to the two nearest planets, Mars and Venus. However mission to Mercury and Jupiter are also of prime interest. The terms Type I and Type II refer to the two possible elliptical interplanetary transfer orbits. For Type I, the heliocentric central transfer angle is less than 180° and for Type II it is greater than 180°. Class I or II refers, to planetary encounter at the first or second intersection of the spacecraft trajectory with the planetary orbit.

1.2 DATA RATE

The high data rates to which this study is addressed are appropriate to requirements for photographic, infrared, or other radiation mapping of the planetary surface. Program goals are 10^6 bits per

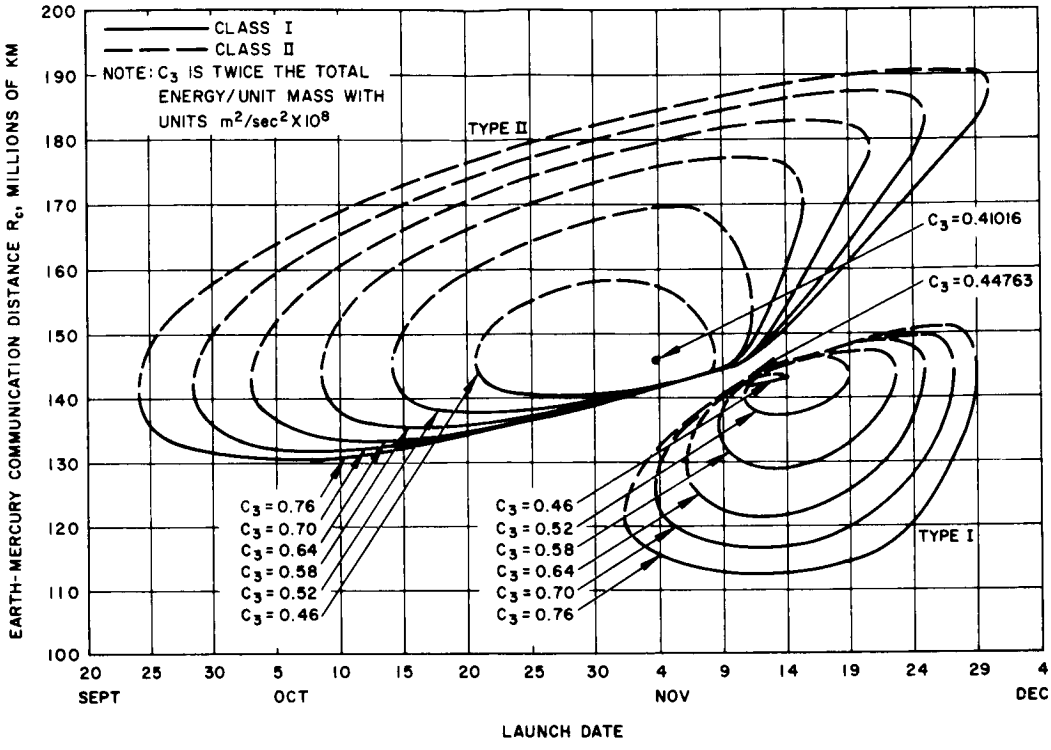


Figure 1-1. Mercury 1968: Earth-Mercury communication distance versus launch date.

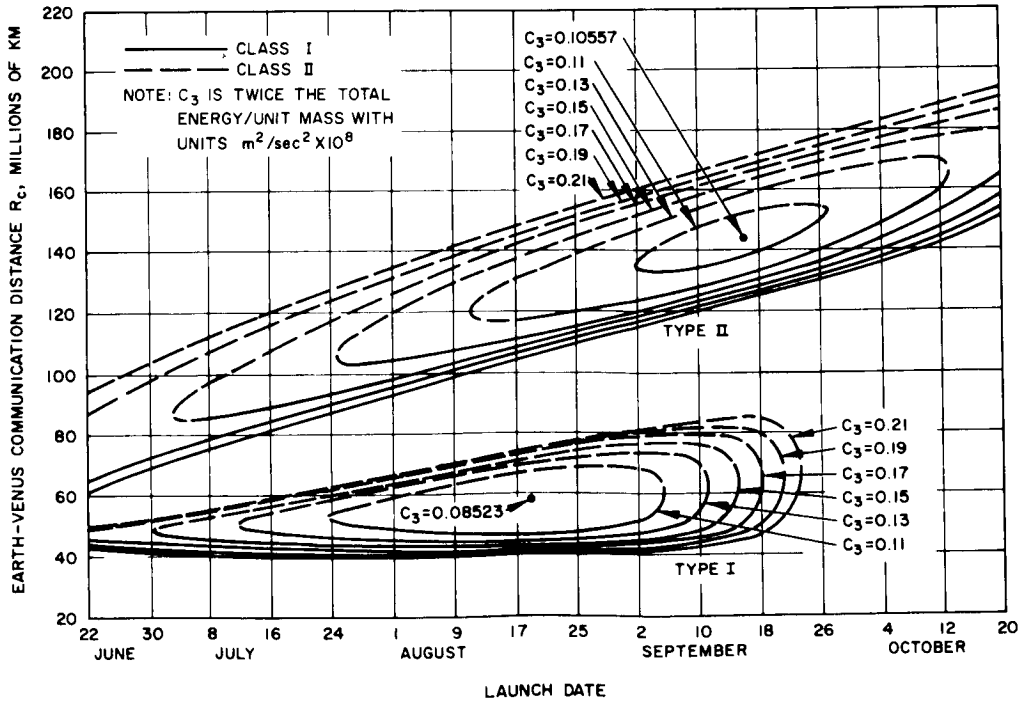


Figure 1-2. Venus 1970: Earth-Venus communication distance versus launch date.

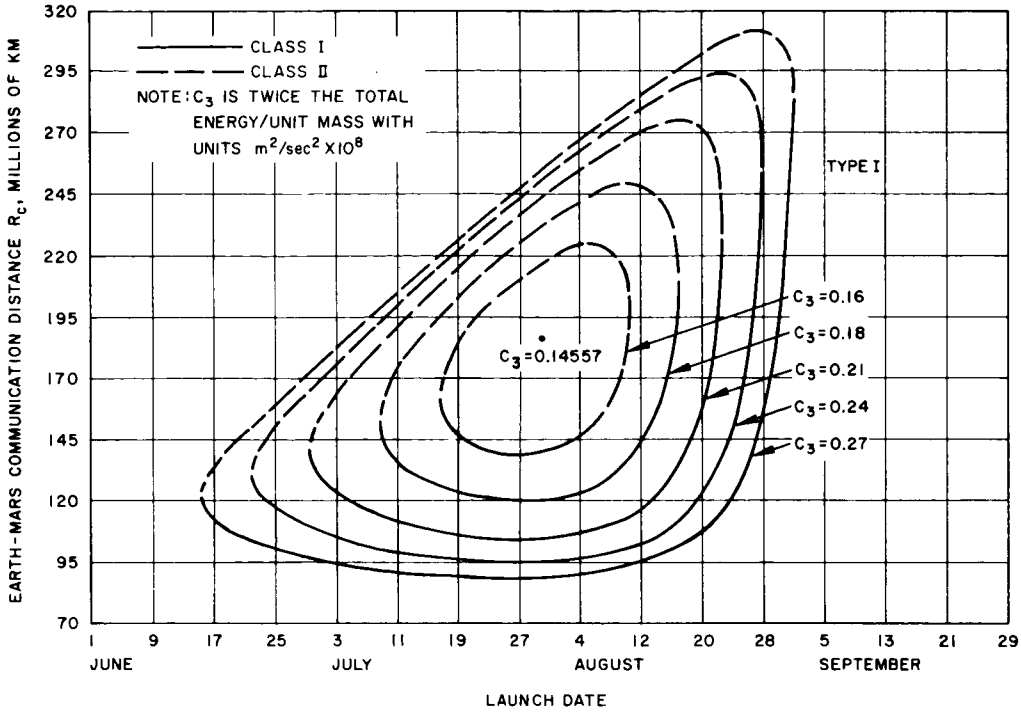


Figure 1-3. Mars 1973: Earth-Mars communication distance versus launch date.

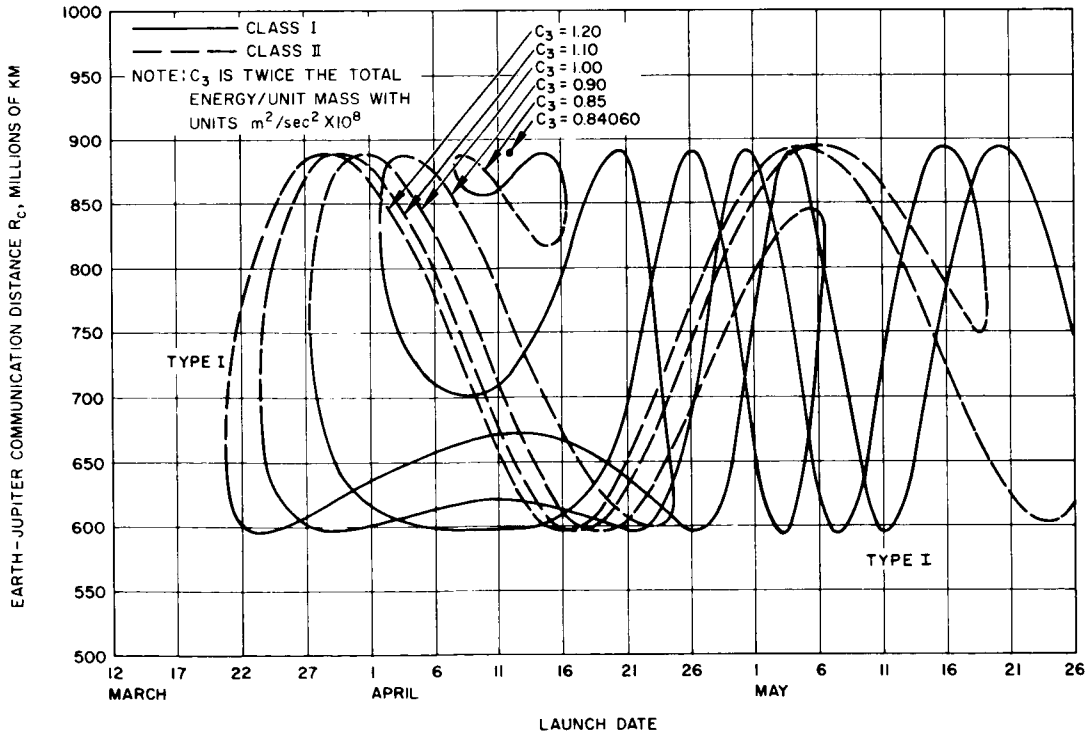


Figure 1-4. Jupiter 1973: Earth-Jupiter communication distance versus launch date.

second for a microwave communication link and 10^8 bits per second in the optical region. The higher rate is adequate for standard quality television where a 6-bit word is used to provide a 64-level gray scale with 4 MHz video sampled twice per cycle. The one megabit rate can provide real time transmission of pictures of comparable quality at a rate of about one every three seconds, or at a faster rate at lower optical or gray-scale resolutions. It has been estimated[4] that photographic reconnaissance of the surface of Mars in sufficient detail to permit landing site selection would comprise of the order of 10^{13} bits of information; this would constitute over 100 days of continuous transmission at the 10^6 bit rate, or just over a day at the 10^8 bit rate. Thus these rates represent a reasonable bracketing of transmission requirements for planetary reconnaissance missions. With reference to the approximate 10-bit-per-second rate for Mariner IV, performance improvements of 50 to 70 db are required to meet the program goals.

Figure 1-5[5] is a projection of communications requirements from present capability for a nominal range of 386 million Km. This figure tends to confirm the goals as logical extensions of present capability.

1.3 ANGULAR COVERAGE

The angular coverage required for spaceborne antennas is dependent on mission requirements and vehicle design. For relatively short communication links between a lander and its orbiting bus, or between a satellite station and Earth, low gain antennas may be appropriate to minimize pointing requirements. For the critical space vehicle to Earth (or satellite) link, high-gain antennas are needed, but the required angular coverage is limited to the ecliptic plane. At microwave frequencies where the beamwidth associated with practical spaceborne-antenna gains considerably exceeds the angle subtended by the Earth, the spacecraft may be oriented so that pointing requirements are restricted to a single degree of freedom. The angular coverage required of the antenna in the ecliptic then will depend on the duration of the mission and other mission requirements which may restrict spacecraft orientation with respect to Earth. At optical

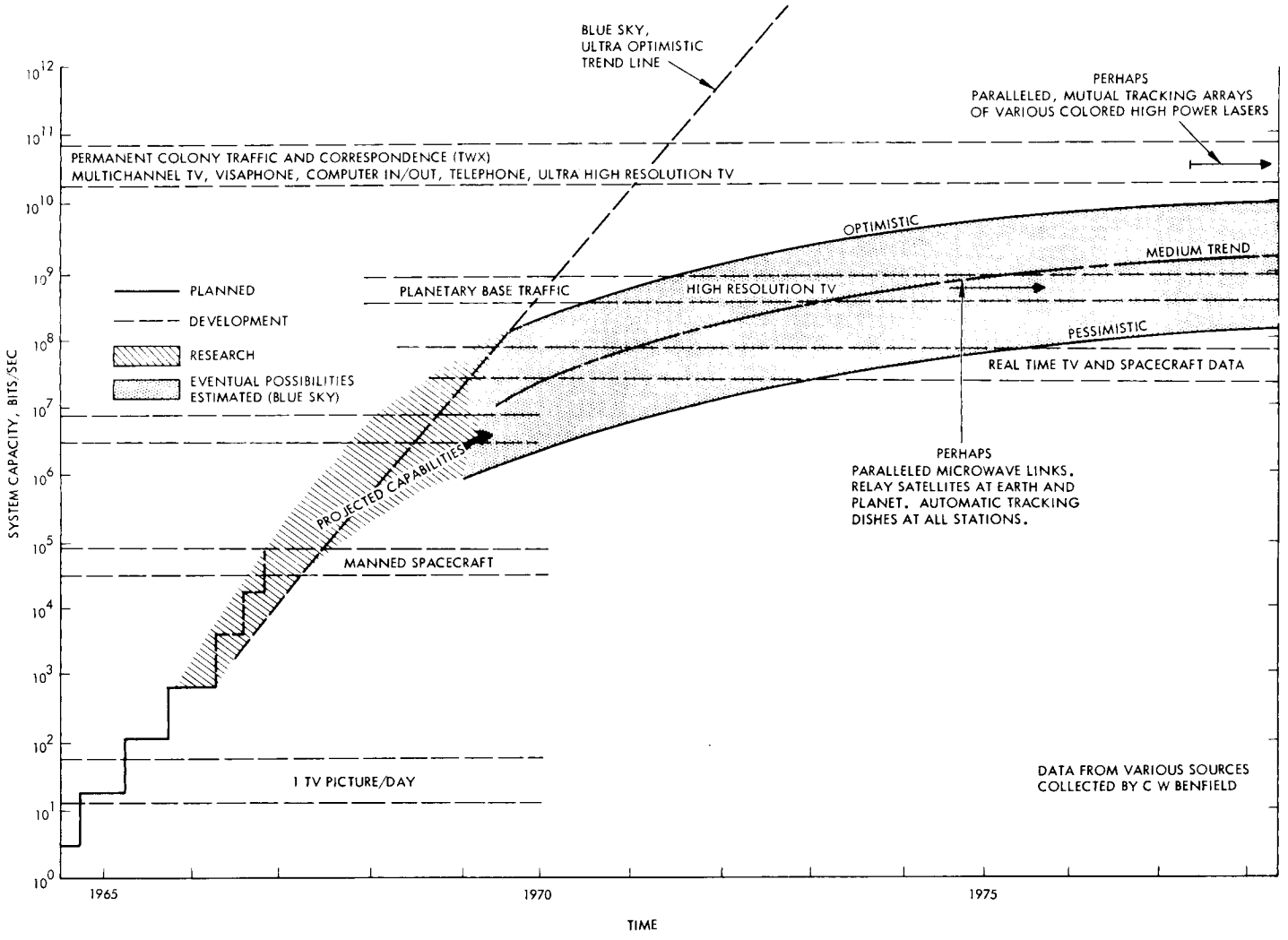


Figure 1-5. Projected communication requirements for near planets. Message time, or data rate, can be improved by technological advances. Data rate appears as bits/second for a sample planetary distance of 386 million kilometers, with mission capabilities represented by various bit rates.

frequencies where beamwidths narrower than the Earth's subtense are to be expected (for example, a 10 microradian beam at a distance of 1 AU subtends a circle of less than 1000 miles diameter), fine tracking of the transmitter beam becomes a much more critical factor.

For a ground-based station in a direct spacecraft-to-Earth link, angular coverage must include, in general, the ancient zodiac, a strip ±8-1/2 degrees about the ecliptic within which lie all of the major planets except Pluto. Since the earth is inclined at an angle of 23-1/2 degrees to the ecliptic, the total declination angle to be covered is ±32 degrees, and ground stations must be located within 50 degrees of the equator. The normal to the antenna aperture typically is normal to the axis of the earth. For M stations evenly spaced in longitude around the earth covering a declination band ±D and a longitudinal arc $2\pi/M$, the maximum beam excursion from the normal is ψ, where

$$\cos \psi = \cos D \cos (\pi/M)$$

Values of ψ are given in Table 1-1 for the maximum declination angle, D = 32 degrees and for various M.

Number of Stations, M	Maximum excursion angle, ψ (degrees)
2	90
3	64.9
4	53.2
5	46.7
6	42.7
.	.
.	.
∞	32

Table 1-1. Maximum angular excursion of beam from normal for M ground stations.

From the table it is evident that three stations represents the practical minimum and there is little incremental improvement for additional stations above four. Four stations may be desirable, however, to limit atmospheric losses and system noise temperature at minimum elevation angles and to minimize losses associated with the scanning of large antenna arrays. For the same reasons the stations should be located as near the equator as possible.

1.4 REFERENCES

1. G. W. Morgenthaler and G. E. Fosdick, Selection of Launch Vehicles, Spacecraft, and Missions for Exploration of the Solar System, Research Report R-64-6, The Martin Company, 1964.

2. Summary Report, Future Programs Task Group, National Aeronautics and Space Administration, Washington D.C., 1965.

3. Design Parameters for Ballistic Interplanetary Trajectories, Jet Propulsion Laboratories Technical Report No. 32-77; Part I, "One Way Transfers to Mars and Venus" January 16, 1963 and Part II, "One Way Transfers to Mercury and Jupiter," January 15, 1966.

4. H. B. Hallock, J. Grusauskas, and D. R. Lamberty, The Potential Role of Optical Communications in the Exploration of Mars, Proc. AIAA/AAS Stepping Stones to Mars Meeting, Baltimore, Md., March 28-30, 1966.

5. C. W. Benfield, NASA Pattern Relevance Guide, Volume IV, Honeywell, Military Products Group, October 1965.

2.0 BASIC SYSTEM LIMITATIONS

The electromagnetic transmission of energy is governed by the one-way transmission equation:

$$S = \frac{G_T G_R P_T \lambda^2}{(4\pi)^2 R^2 L}$$

where

S = signal power received

G_T = transmitter antenna gain

G_R = receiver antenna gain

P_T = transmitter power

λ = signal wavelength

R = transmission pathlength

and

L = transmission losses (greater than unity)

The signal power required at the receiver, however, is determined by the required data accuracy and by the noise present, due both to external sources and to the receiver itself. Thus noise presents a fundamental limitation on system performance and can be accounted for by rewriting the transmission equation in terms of the signal-to-noise ratio,

$$S/N = \frac{G_T G_R P_T \lambda^2}{(4\pi)^2 R^2 L hf \left[1 + (e^{hf/kT} - 1)^{-1}\right] B}$$

where

h = Planck's constant

f = c/λ = signal frequency

k = Boltzmann's constant

T = effective absolute temperature of the receiver

and

B = receiver bandwidth

Restrictions on the minimum usable values for the signal-to-noise ratio are discussed in Section 2.1.

The expression $hf\left[1 + (e^{hf/kT} - 1)^{-1}\right] B$ represents the ideal noise limit[1] where T is taken as the temperature faced by the ideal receiver input. In the microwave region where $kT \gg hf$, this expression converges to the familiar quantity kTB, and in the optical region where generally $kT \ll hf$, to hfB. For non-ideal systems detection efficiency and the additional noise contribution due both to external and internal noise sources can be included by taking T as the equivalent system noise input temperature of the receiver. In the optical region where the temperature term does not apply, detector efficiency can be separately accounted for by taking the ratio of effective to ideal input noise to be inversely proportional to detector quantum efficiency. Sources of external noise are discussed in Sections 2.2 and 2.3.

The signal losses, L, included in the transmission equation comprise small losses due to inefficiencies of transmitting and receiving antennas, feeds, etc., and transmission losses due to the fundamental propagation characteristics of the earth's atmosphere. The signal losses due to atmospheric attenuation are discussed as a function of frequency in Section 2.4. Losses due to plasma effects during atmospheric entry by a high velocity vehicle are discussed separately in Section 6.

The effective gains, G_T and G_R, of transmitter and receiver antennas, respectively, may be limited by atmospheric propagation effects as well as by practical limitations on achievable fabrication tolerances. Wavefront distortions due to atmospheric inhomogeneities across the aperture will have an effect similar to that caused by deviations in the antenna surface. Thus atmospheric distortion can set a basic limit to the effective gain for a single aperture. Distortion characteristics of the atmosphere as they are presently known are discussed in Section 2.5.

2.1 SIGNAL-TO-NOISE AND BANDWIDTH LIMITATIONS

The minimum allowable signal-to-noise ratio and signal bandwidth requirement for a given data rate $\mathcal{R}$ in bits per second are determined by the information coding method employed. Based on Shannon's work, the limiting data rate in terms of signal-to-noise ratio and bandwidth is given by the expression:[2]

$$\mathcal{R} \leq B \log_2 (1 + S/N)$$

or

$$\mathcal{R} \leq B \log_2 (1 + R_o/B)$$

where $(S/N)\,B$ is defined as the information-rate parameter, $\mathcal{R}_o$. The maximum data rate can be approached with negligible error by a proper choice of coding technique.[3,4,5] A simple and fairly efficient technique, for example, is coherent biphase coding. The characteristics of this code in terms of signal-to-noise and bandwidth-to-data-rate ratios, and its relation to the Shannon limit are shown in Figure 2-1. For small error probabilities the bandwidth required is comparable with the data rate, an increase in signal serving to reduce the error probability without appreciable effect on the bandwidth requirement. Tolerable values of error probability range from 10^{-5} to 10^{-2} depending on type of data.[6] Thus the practical limit for the product of signal-to-noise ratio and bandwidth, even with a simple code, need

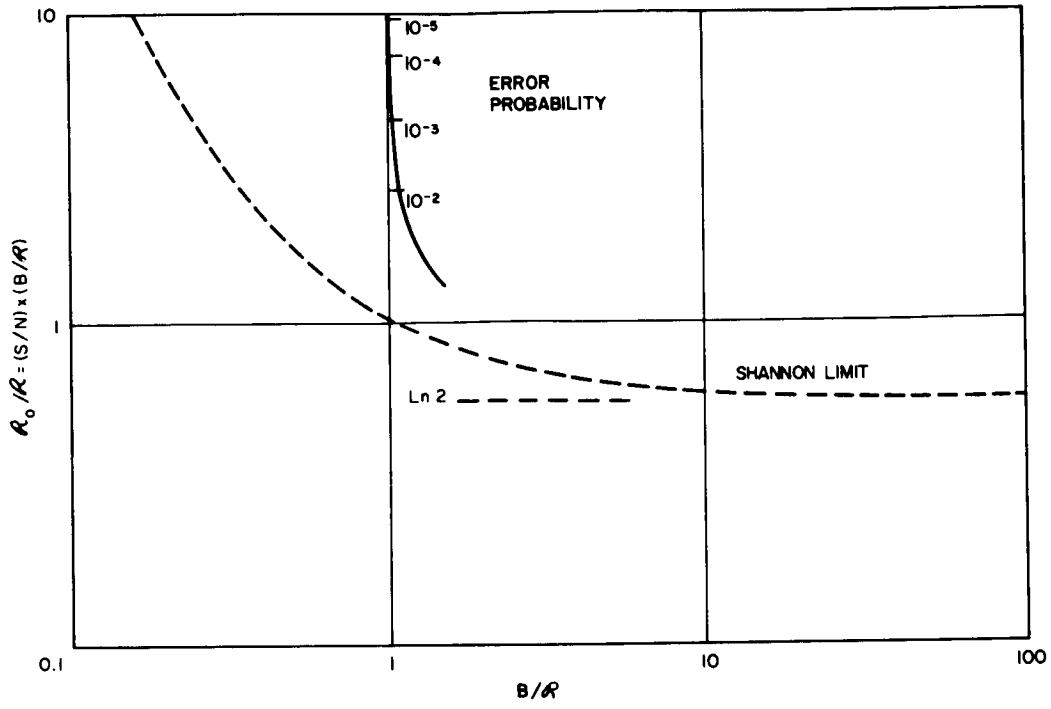


Figure 2-1. Efficiency of biphase coding.

not exceed the ideal limit by more than an order of magnitude to provide acceptable performance. The expression R_o = (S/N) B = 10 will therefore be taken to represent a practical relationship between signal and bandwidth and the limiting noise level. (The actual relationship for a specific system design will depend on the particular coding scheme adopted as well as on error-rate requirements.)

2.2 RADIO FREQUENCY NOISE SOURCES

Relative contributions to microwave system noise from various sources may be conveniently compared in terms of antenna noise temperature. If the noise power incident on the antenna is constant over the bandwidth of interest, the antenna temperature is

$$T_a = \frac{P}{kB}$$

where

T_a = antenna temperature (^oK)

P = incident noise power (watts) in the plane of polarization of the antenna

B = bandwidth (Hz)

k = Boltzmann's constant (1.38×10^{-23} joules/^oK)

Noise from extended sources, large relative to the antenna beamwidth, is commonly specified in terms of brightness or brightness temperature. Brightness is defined as irradiance in watts/meter^2 per Hz of bandwidth incident on an antenna having an angular beamwidth of one steradian. The brightness temperature of an extended source at a given frequency is defined as the temperature of a blackbody surface which would produce the same brightness at that frequency. The brightness temperature is related to the brightness by the Rayleigh-Jean's approximation to Planck's law,

$$T_b = \frac{\beta\lambda^2}{2k}$$

where

T_b = brightness temperature (^oK)

β = brightness (watts/meter^2 steradian) per Hz

λ = wavelength (meters)

The antenna temperature due to an extended source is then

$$T_a = T_b = \frac{\pi\beta\theta_a{}^2 A_a}{8k} = \frac{2\pi\beta A_a}{kG}$$

in terms of brightness where θ_a is the angular beamwidth of the antenna and is small, A_a is the effective antenna area, and G, the gain.

Noise from sources having small angular subtense (θ_s) relative to the antenna beamwidth is customarily specified in terms of irradiance (watts/meter^2)/Hz of bandwidth. Since the irradiance, H, is just $H = \beta\ (\pi/4)\ \theta_s^2$, the source temperature in accordance with the Rayleigh-Jean's approximation to Planck's law is

$$T_s = \frac{2\lambda^2 H}{\pi k\ \theta_s^2}$$

The antenna temperature is given by

$$T_a = \frac{H\ A_a}{2k} = \frac{2H\lambda^2}{\pi k \theta_a^2}$$

in terms of irradiance, or by

$$T_a = \left(\frac{\theta_s}{\theta_a}\right)^2 T_s$$

in terms of source temperature.

2.2.1 Galactic Background

The sun is a small star in one arm of our spiral planar galaxy. Antennas see a higher temperature source when pointed along this galactic plane than otherwise. Eight of the most reliable measurements of galactic noise spatial distributions have been summarized[7] and give maximum and minimum backgrounds as shown in Figure 2-2.[8] The curves of maximum galactic background versus frequency represent the brightness temperature at the galactic center while the minimum curve is for one of the coldest regions of the radio sky. It is evident that the galactic background can be neglected for frequencies above 2 Ghz.

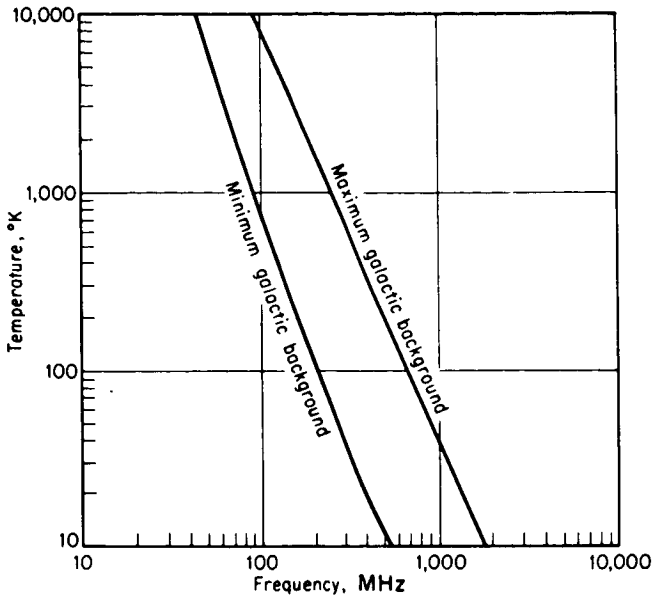


Figure 2-2. Galactic noise temperature.

2.2.2 Radio Stars

Superimposed on the general galactic background radiation are numerous discrete sources, generally less than 1 degree in extent. Since the majority of these sources cannot be identified with visible objects they are known as radio stars. The strongest of these sources tend to occur near the plane of the galaxy. Figure 2-3[8] locates the ten strongest radio stars in equatorial coordinates and gives their noise temperatures at 378 Mhz. The spectra of the four brightest radio stars are shown in Figure 2-4.[9]

2.2.3 Hydrogen Line Emission by Interstellar Gas Clouds

Radio astronomers have detected an essentially monochromatic spectrum line due to the radiation of atomic hydrogen at a wavelength of 21 cm (1.42 Ghz). The hydrogen line emission is a maximum along the Milky Way with a distribution over the sky roughly similar to the general galactic background. It has a maximum brightness temperature of 100° and maximum brightness of 6×10^{-20} watts/m^2 steradian per Hz[10]. This is well in excess of the general galactic background at that frequency. However, since this is limited to a single wavelength, it is unimportant as a source of interference.

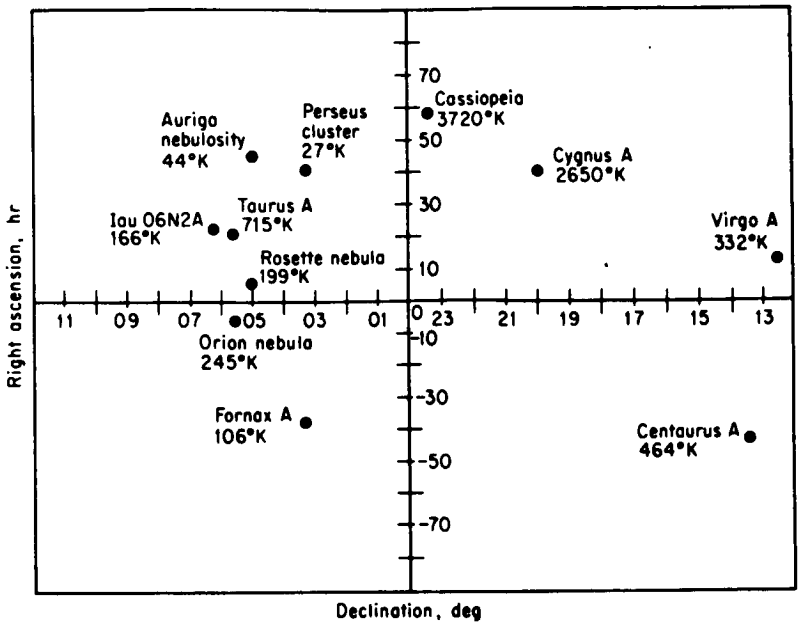


Figure 2-3. Radio stars at 378 MHz.

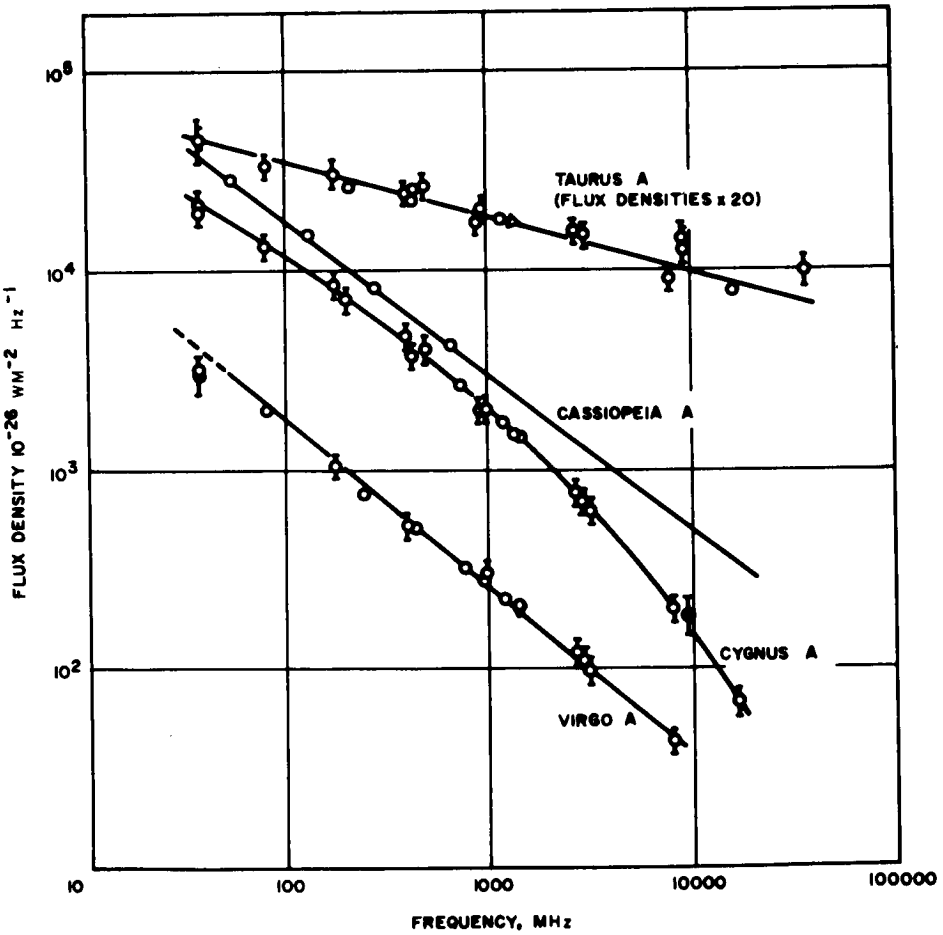


Figure 2-4. Spectra of the most prominent radio stars.

2.2.4 Solar Radiation

The equivalent noise temperature of the quiet sun varies from 7×10^5°K at 0.3 GHz to 6000°K at 30 GHz. The observed values of sun temperature between 0.25 and 35 GHz follow closely the relationship[11]

$$\frac{T_s}{290} = \frac{675}{f}\left[1 + \frac{1}{2.3}\sin 2\pi \frac{\log_{10} 6\,(f - 0.1)}{2.3}\right]$$

where T_s is the apparent sun temperature and f is the frequency in gigahertz. This equation is plotted in Figure 2-5.

From earth, the sun subtends a solid angle of about 7×10^{-5} steradians. The temperature given by the above equation will be observed by an antenna having a beamwidth equal to or less than this. For antennas having beamwidth greater than this, the apparent temperature is decreased by the ratio of the sun angular subtense to

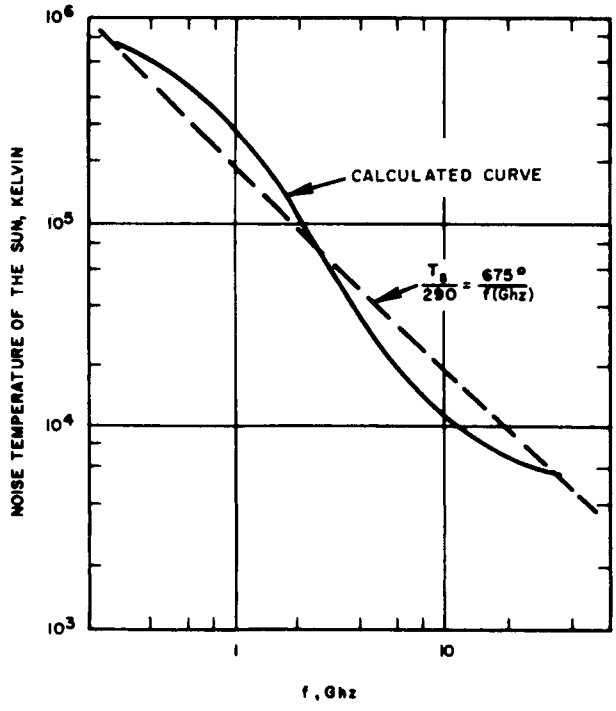


Figure 2-5. Sun noise temperature versus frequency.

the antenna beamwidth (solid angle). Measured solar irradiance levels outside the earth's atmosphere are indicated in Figure 2-9.[17]

2.2.5 Planetary Radiation

For low noise, high gain, receiving systems, planetary radiation can become significant (see for example Table 4-11). The planets produce radio frequency radiation by thermal emission. At radio wavelengths, thermal radiation intensity varies as $1/\lambda^2$ (Rayleigh-Jean's approximation to blackbody radiation at long wavelengths).

Figure 2-6[12] shows the theoretical maximum thermal radiation intensity (irradiance) outside the earth's atmosphere based on planetary temperature estimates. These are essentially in agreement with infrared measurements except in the case of Jupiter, Venus, and Saturn. Actual measured radiation from these planets is included in Figure 2-6.

In addition to this thermal radiation, strong bursts of lower frequency impulsive radiation have been observed from Jupiter and Venus. This non-thermal radiation has not been observed at frequencies greater than 43 MHz.

2.2.6 Lunar Radiation

In the microwave region, the moon acts as a blackbody radiator at a temperature which lags by about 45° in lunar phase that determined by infrared measurements. While lunar surface temperatures as determined by infrared measurements vary from 120°K to 400°K as the lunar phase changes, brightness temperatures at 35 GHz[13] vary from 145°K to 220°K. At frequencies less than 1.4 GHz the radiation temperature is approximately constant at 250°K.[8]

2.2.7 Terrestrial Background

Thermal radiation from the ionosphere is very small for microwave frequencies. The effect of ionospheric radiation can be approximated by assuming that each 0.1 db of absorption is equivalent to 7°K antenna noise temperature. Since the ionospheric absorption is less

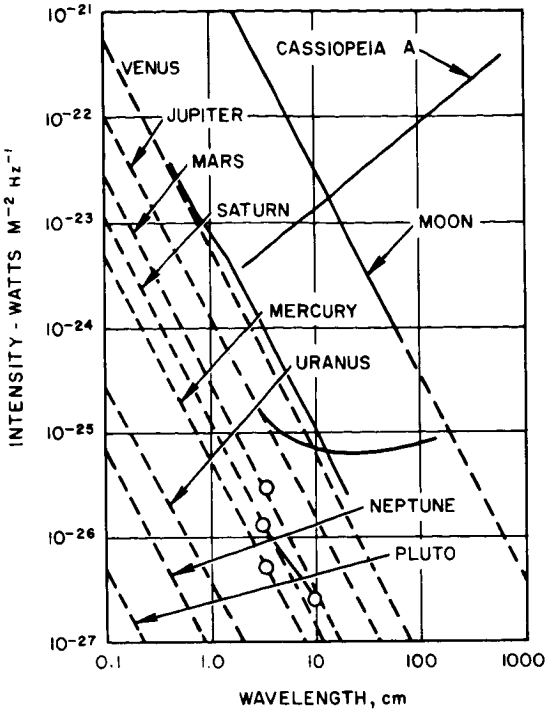


Figure 2-6. The predicted intensity of planetary thermal radio radiation when the planet is near closest approach to the earth (dashed lines), and the measured intensity of radiation from the planets, the moon, and the nonthermal radio source, Cassiopeia A (solid lines and points).

than 0.1 db for frequencies greater than 0.3 GHz,[14] this noise source can be neglected in comparison with the galactic background.

In the troposphere, thermal radiation is primarily due to oxygen and water vapor. The resultant tropospheric contribution to the noise temperature of a narrow beam antenna whose radiation pattern has no side or back lobes has been calculated for various zenith angles and is shown in Figure 2-7.[11] The calculated curves are in essential agreement with experimental measurements.

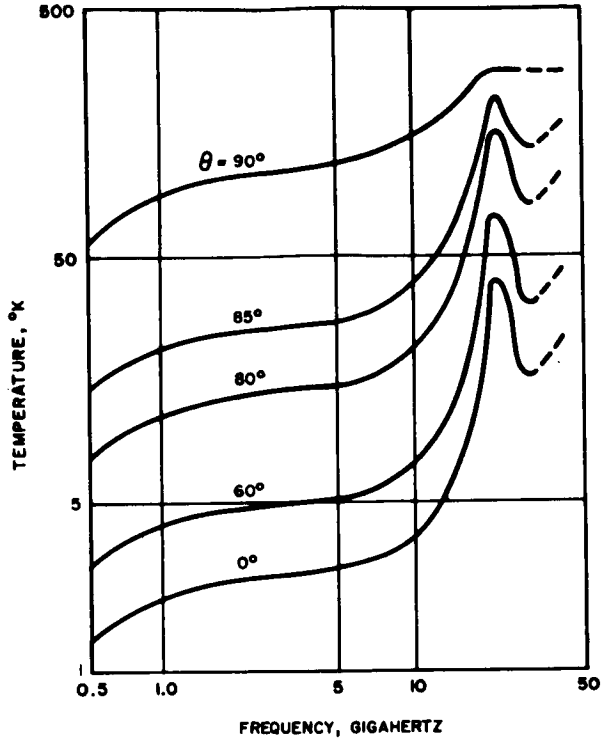


Figure 2-7. Sky noise temperature due to oxygen and water vapor at various zenith angles, (θ).

Thermal radiation from the earth itself can be approximated by blackbody radiation at 300°K in the microwave and millimeter wave region of the spectrum. This is an important source since approximately 2π steradians of back lobe angular coverage is illuminated by it. Thus low back lobes are an important consideration in low noise earth based antennas.

2.3 OPTICAL NOISE SOURCES

For an optical receiver with a very narrow field of view such as is appropriate to a deep space communication link, background noise will often be negligible. With a 10-microradian field of view, for example, the effective irradiance from a daytime sky is of the order of the irradiance from the brightest stars. The effective irradiance of the moon and the nearer planets is reduced to a similar level

inasmuch as only a portion of their surfaces will be within the field of view. The maximum background irradiance therefore will be of the order of 10^{-11} watts/cm^2 micron at a wavelength of 0.5 micron. At a bit rate of 10^8 bits/second this corresponds to less than one photon per bit for a receiver having an aperture area of one square meter and an optical bandwidth of one Angstrom. Thus, except for a direct view of the sun, optical receivers will generally be limited by signal statistics or detector noise. This will be particularly true of coherent receivers where optical noise bandwidths correspond more nearly to signal bandwidths.

However, for receivers having relatively wide fields of view or very large apertures, such as may be appropriate to non-coherent systems, noise contributions from the external sources discussed below may be significant.

2.3.1 Stellar Radiation

The spectral irradiance of the stars has been calculated assuming blackbody radiation.[15] Figure 2-8 depicts spectral irradiance reaching the top of the earth's atmosphere from the stars which exhibit the largest irradiance in the visible region. Variable stars such as Betelguex, Mira, and R. Hydrae are described by their maximum irradiance.

2.3.2 Solar Radiation

The spectral distribution of solar radiation resembles that of a blackbody at 6000°K. Over 98 percent of the solar radiation is contained in the wavelength region 0.3 to 4.0 microns. Figure 2-9[16] shows the spectral distribution of solar energy just outside the atmosphere at the earth's mean solar distance (1 AU).

2.3.3 Lunar and Planetary Background Radiation

The spectral irradiance of the moon and the brighter planets due to reflected solar radiation and self emission is depicted in Figures 2-10 and 2-11.[6] The spectral self emission has been

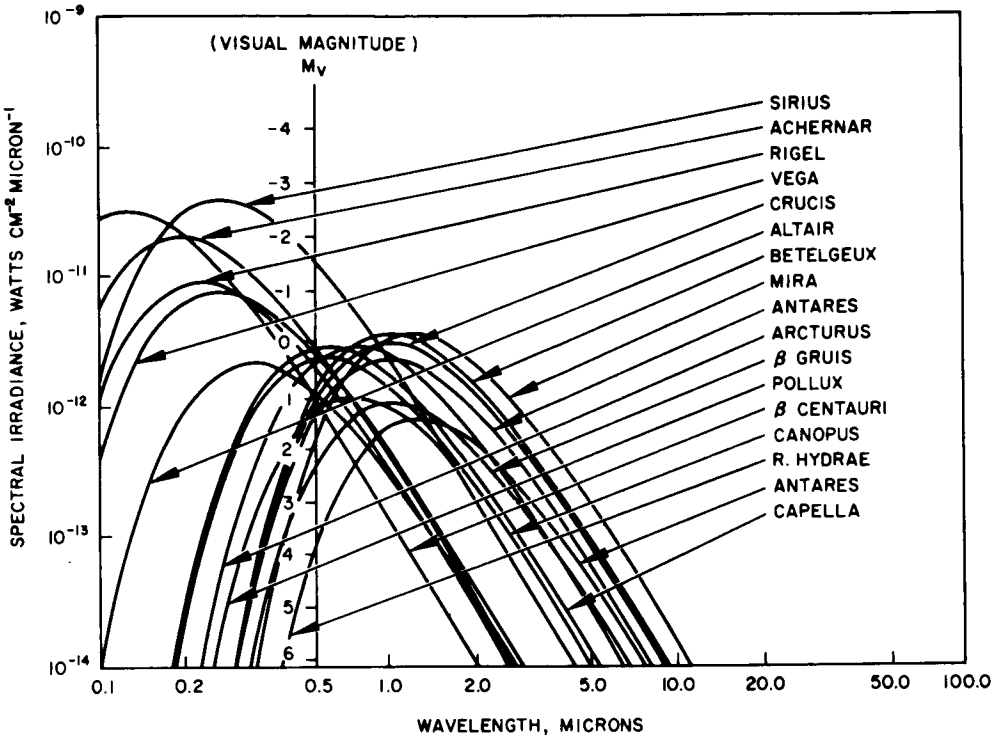


Figure 2-8. Spectral irradiance of brightest stars outside the terrestrial atmosphere.

calculated on the assumption that the planets radiate as grey bodies. Irradiance from the planets Neptune and Pluto reaching earth is insignificant.

Earth radiation to space consists of reflected solar energy and self-emitted energy. For reflected energy the mean albedo over the solar spectrum is approximately 0.35. In addition to this the earth radiates by self-emission as a blackbody at a temperature of 220°K to 320°K, depending on latitude. When the sum of the solar reflected and blackbody radiation from the earth's surface is combined with the effect of selective atmospheric absorption, the spectral radiant emittance of Figure 2-12[17] results.

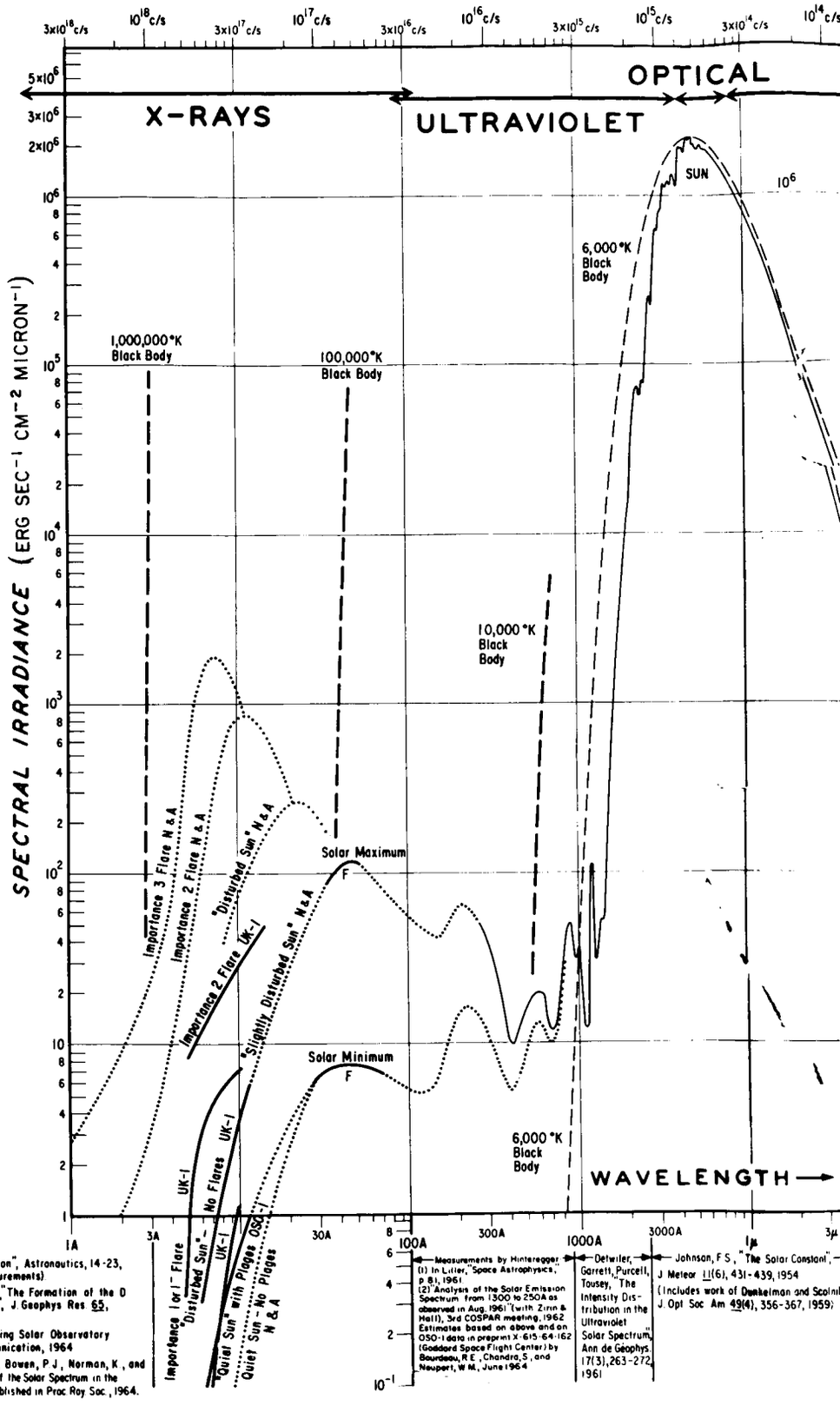
THE SOLAR SPECTRUM

Displayed here is the entire radiation pattern of the sun, from X-rays through visible light to radio waves. The solar energy received at the top of the earth's atmosphere can be read from the vertical scales. Wavelengths to which these energies apply are indicated at the bottom, frequencies at the top.

The units of spectral irradiance read, in unabbreviated form, "ergs per square centimeter per second in a wavelength interval of one micron." Because of the enormous range of energies charted here, it was necessary to fold this logarithmic scale thrice in the infrared and radio regions. The folded sections are accompanied by the appropriate numerical labels.

Along the scale at the top, frequencies in cycles per second (c/s) increase toward the right. One gigacycle per second (Gc/s) is equivalent to 10^9 (one billion) cycles per second; therefore, 1,000 gigacycles equals 10^{12} cycles. Farther to the right, one megacycle per second (Mc/s) is 10^6 (one million) cycles per second.

The wavelengths, given along the bottom scale, become longer toward the right; they are indicated successively in angstroms (A), microns (μ), millimeters (mm), centimeters (cm), and meters (m). One micron is 10,000 angstroms, and one millimeter is 1,000 microns.



F = Friedman, H., "Solar Radiation", Astronautics, 14-23, August 1962 (rocket measurements)

N & A = Nicolet, M and Aikin, A C., "The Formation of the D Region in the Ionosphere", J Geophys Res 65, 1469-1483, 1960

OSO-1 = Data from the first Orbiting Solar Observatory W A White, private communication, 1964

UK-1 = Pounds, K A., Willmore, A P, Bowen, P J., Norman, K., and Sanford, P W., "Measurements of the Solar Spectrum in the Wavelength Band 4-14A" Published in Proc Roy Soc., 1964. Ariel I satellite data.

Measurements by Hinteregger (1) In Litter, "Space Astrophysics," p 81, 1961 (2) "Analysis of the Solar Emission Spectrum from 1300 to 250A as observed in Aug, 1961" (with Zirin & Hall), 3rd COSPAR meeting, 1962 Estimates based on above and on OSO-1 data in preprint X-615-64-162 (Goddard Space Flight Center) by Bourdeau, R E., Chandra, S., and Neupert, W M., June 1964

Detwiler, Garrett, Purcell, Tousey, "The Intensity Distribution in the Ultraviolet Solar Spectrum," Ann de Geophys 17(3), 263-272, 1961

Johnson, F S., "The Solar Constant", J Meteor 11(6), 431-439, 1954 (Includes work of Dunkelman and Scolnik) J Opt Soc Am 49(4), 356-367, 1959)

2-15

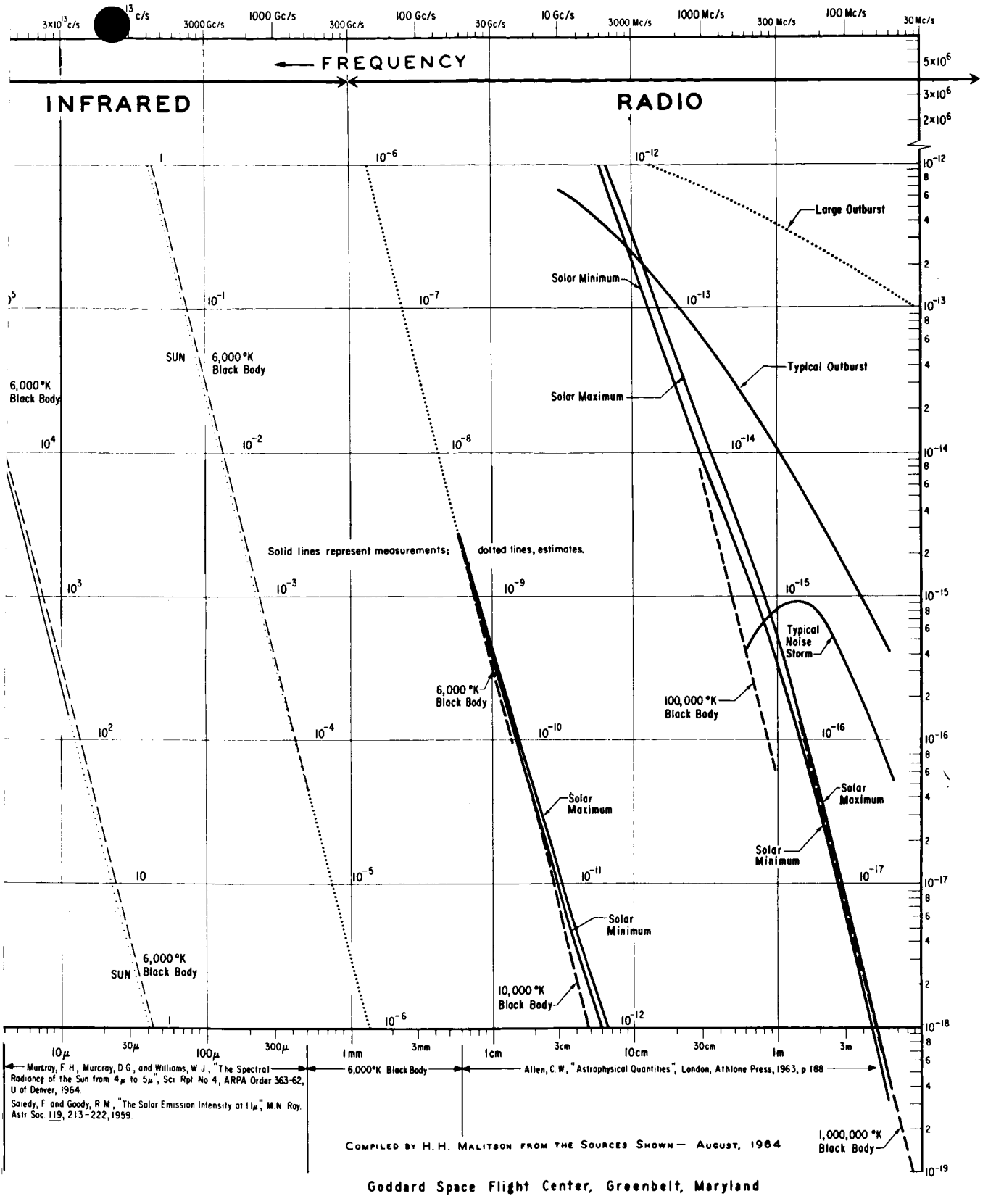


Figure 2-9. The solar spectrum.

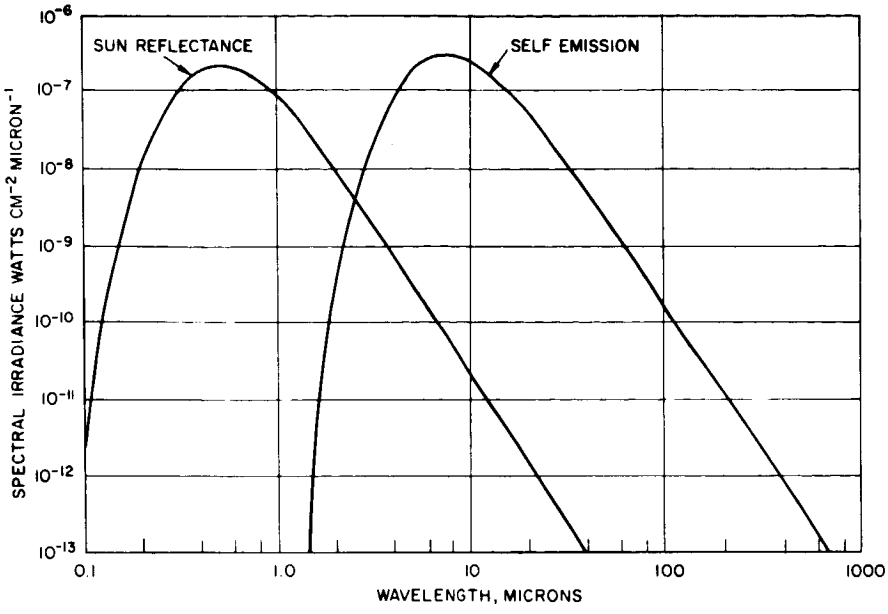


Figure 2-10. Full moon spectral irradiance outside the terrestrial atmosphere.

2.3.4 Terrestrial Atmospheric Background

The radiance of the sky is the result of two mechanisms: molecular scattering of incident radiation and emission by atmospheric constituents as a result of absorption of incident radiation. Atmospheric emission is significant only at wavelengths longer than 2μ. Scattering of solar radiation is the overwhelming contribution to daytime sky radiance in the visible and near visible ragion. Spectral radiance versus wavelength of the clear daytime sky is plotted in Figure 2-13[18] for a 45 degree zenith angle with the sun near zenith. Radiances of sunlit clouds are typically an order of magnitude greater. Measurements of the spectrum of diffuse night radiance[19] give a spectral radiance in the visible and near infrared of the order of 10^{-10} to 10^{-9} watts/cm^2 micron steradian. Except for the narrow intense Na and H atomic lines, relatively light sky emissions appear between 0.1 and 1.0 micron. Beginning at 1 micron, intense OH

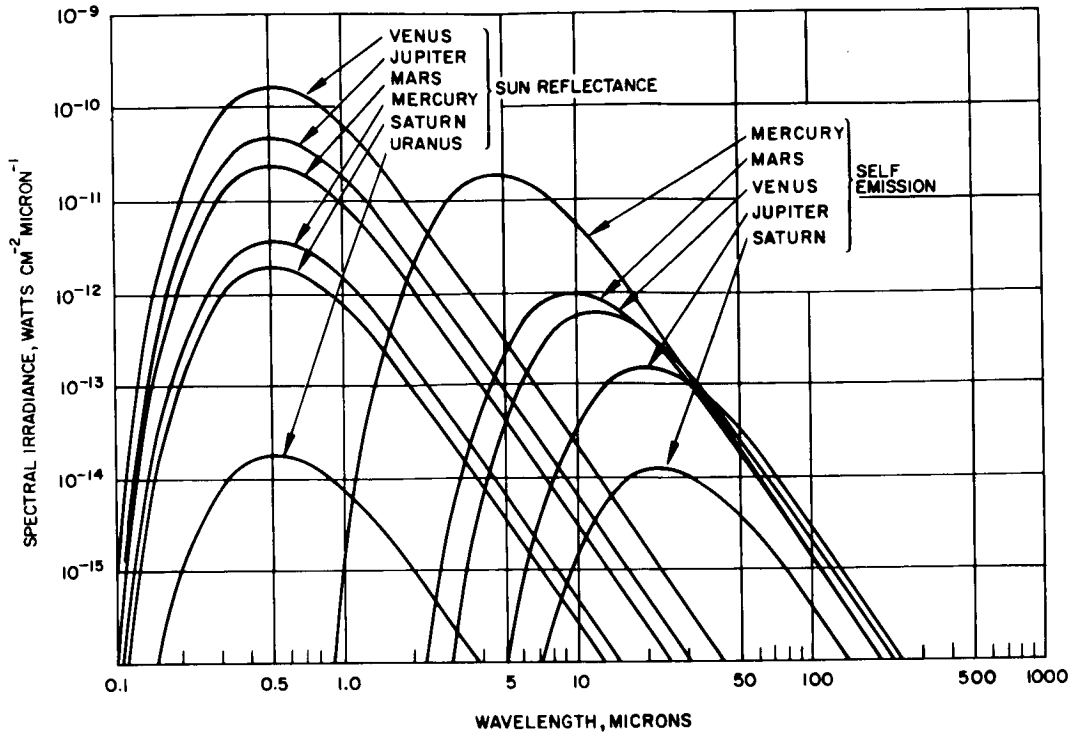


Figure 2-11. Calculated planetary spectral irradiance outside the terrestrial atmosphere.

molecular bands appear as "air glow" with radiances between 10^{-9} and 10^{-8} watts/cm^2 micron steradian. Above 2 microns thermal emission from the dense lower atmosphere corresponding roughly to 260°K blackbody radiation obscures the air glow.

2.4 ATMOSPHERIC ATTENUATION

2.4.1 Attenuation at Radio Frequencies

Ionospheric attenuation becomes negligible above about 300 megahertz[14] so that for microwave transmission through the atmosphere attenuation results principally from absorption due to oxygen, water vapor and water in the troposphere. Theoretical attenuation spectra for these components are shown in Figure 2-14 together with experimental values for several conditions of rain and fog.[20] The expected attenuation for a one-way passage through a standard summer

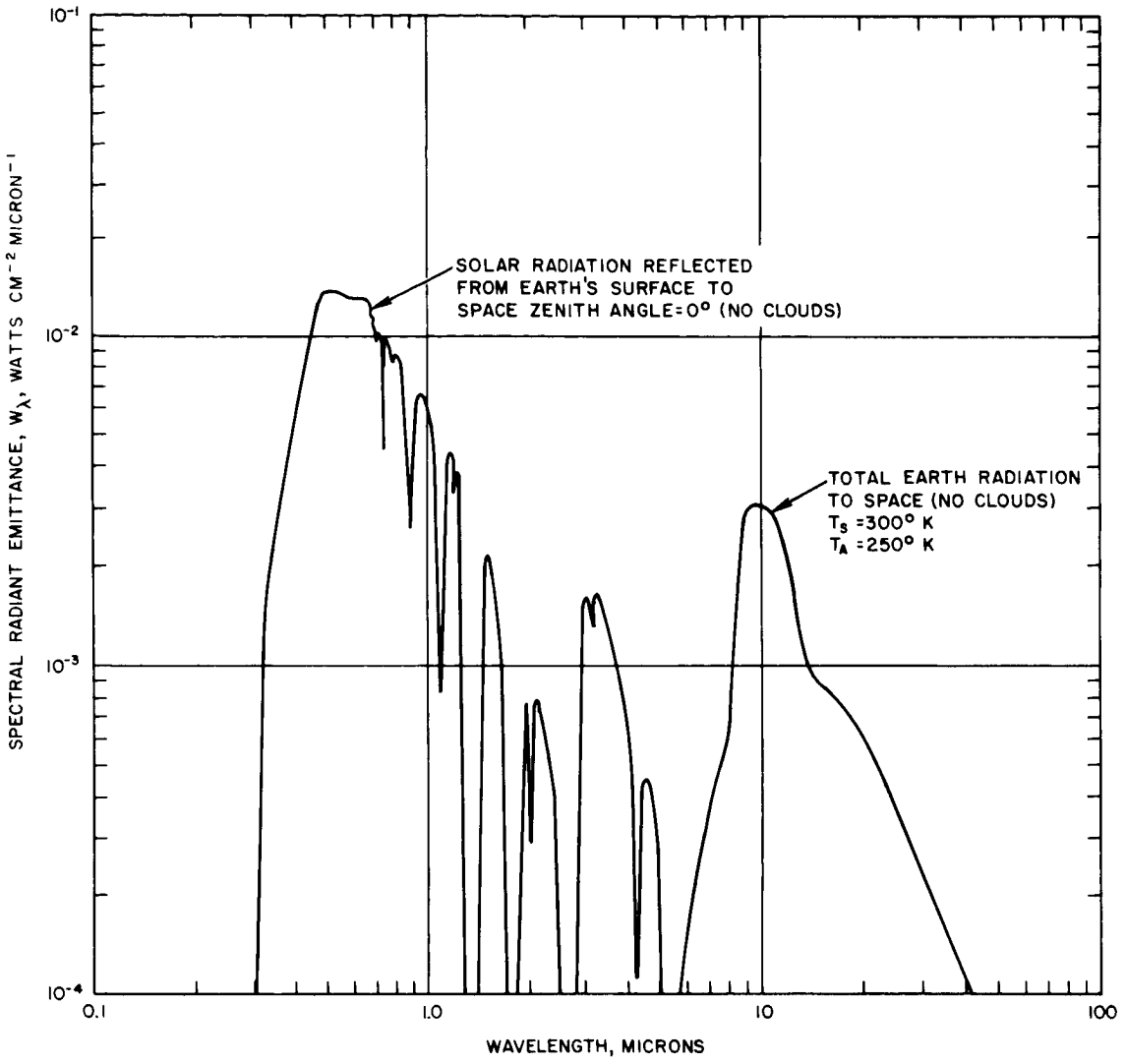


Figure 2-12. Solar and terrestrial radiation. Reflected solar and total earth radiation to space values should be divided by π to obtain the radiance for each case. T_S is the surface temperature and T_A is the effective radiating air temperature.

atmosphere at several zenith angles is given in Figure 2-15.[11] These curves have been derived from calculations made for a two-way radar transit and are in good agreement with experimental data.[21] They indicate that attenuation increases rapidly above 10 gigaHertz.

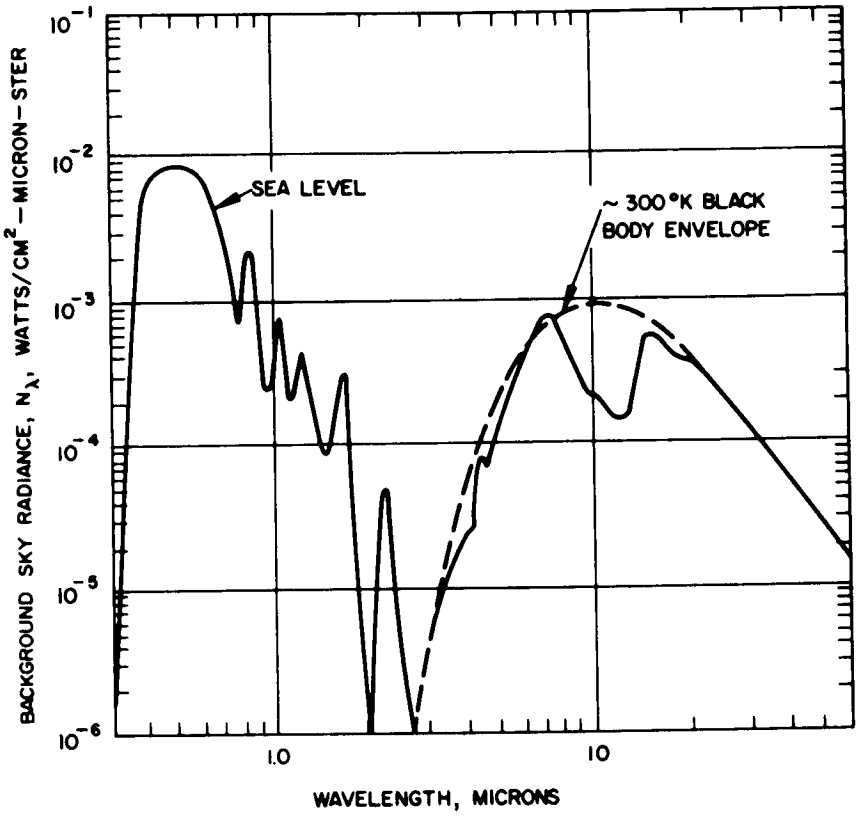


Figure 2-13. Diffuse component of typical daytime sky background at sea level, zenith angle 45°, excellent visibility.

Data are not generally available for the wavelength region between 1 millimeter and 20 microns but it is believed to be blocked for the most part by water vapor and other molecular absorption bands.

2.4.2 Attenuation at Optical Frequencies

There are several transmission windows in the infrared as indicated in Figures 2-16 and 2-17[22] but roughly half of the spectrum is still blocked by molecular absorption bands. The density of absorption bands decreases in the near infrared and visible regions as shown in Figure 2-18[22]. The curves, given for several values of zenith angle, are for very clear atmospheric conditions.

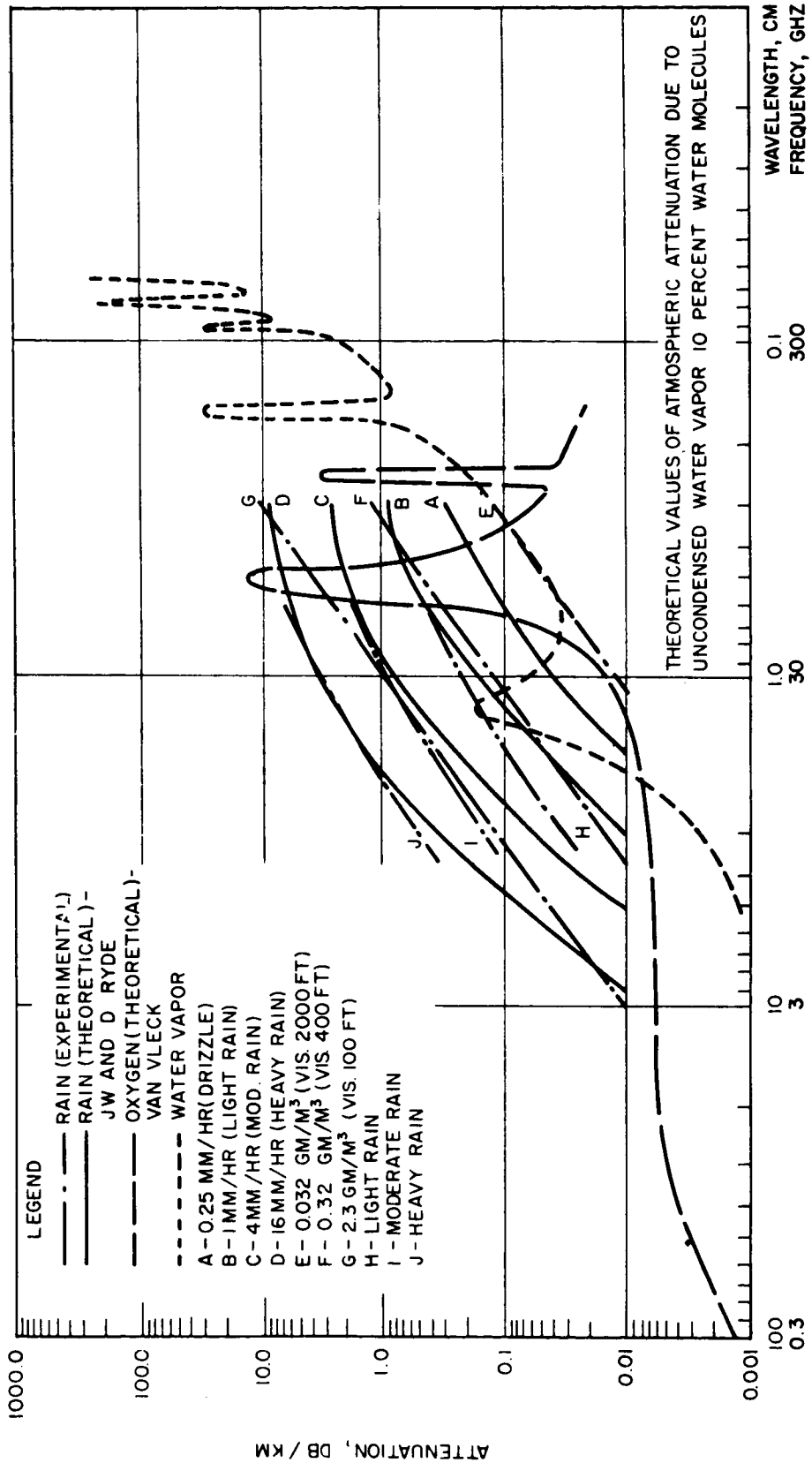


Figure 2-14. Atmospheric attenuation summary.

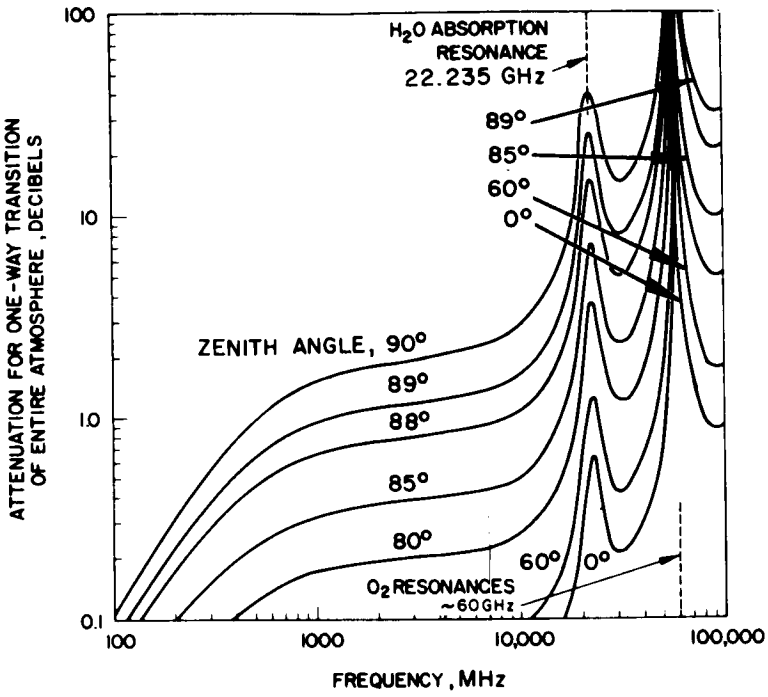


Figure 2-15. One way attenuation through the standard summer atmosphere due to oxygen and water vapor.

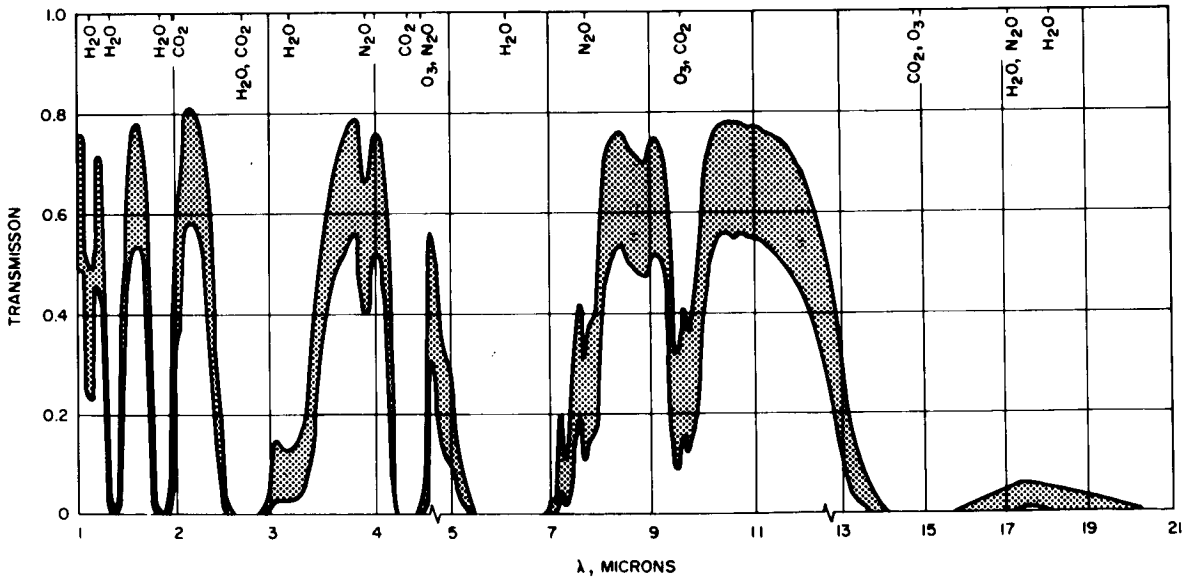


Figure 2-16. Transmission of the atmosphere in the infrared at sea level for zenith angles from 20 to 70 degrees. Absorbing factors: CO_2, N_2O, CH_4, CO, O_3, and haze (visibility = 32 km).

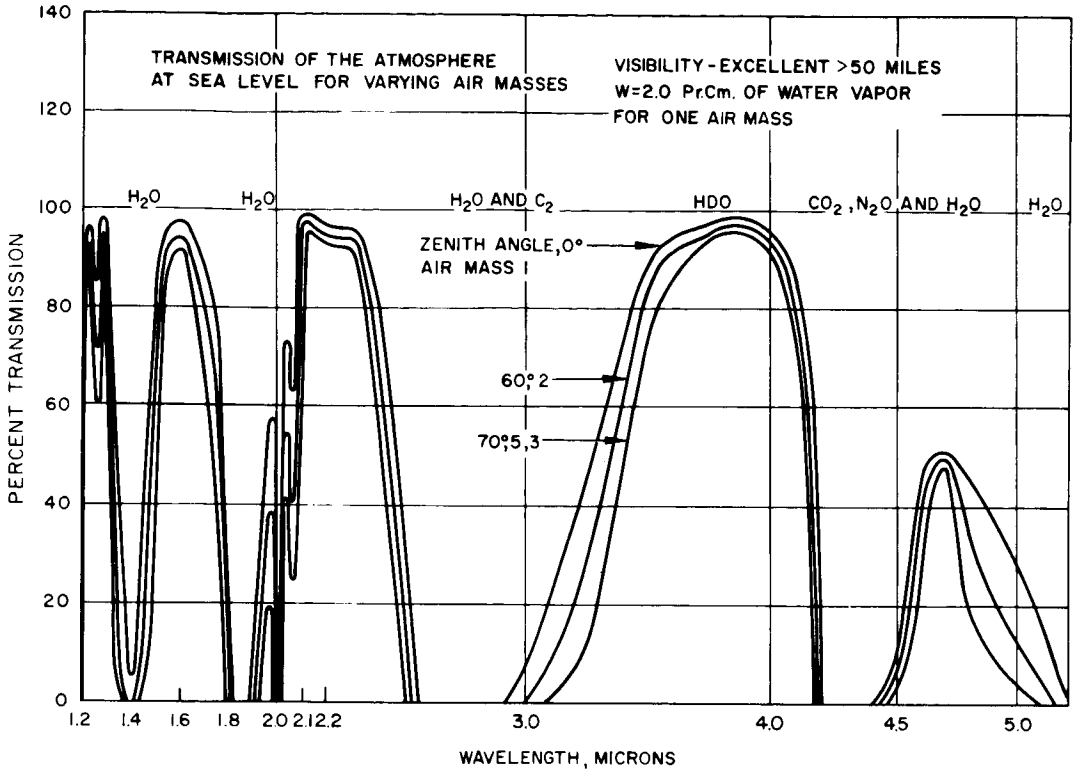


Figure 2-17. Transmission of the atmosphere at sea level for several zenith angles, 1.2 to 5 microns.

It is important to note that the atmospheric absorption bands comprise a large number of sharp absorption lines not resolved on the scale of the curves shown. For the essentially monochromatic radiation generated by lasers, windows may exist within these bands or conversely, relatively isolated absorption lines may exist in apparent windows. Thus high-resolution spectral measurements are necessary in the vicinity of laser lines of interest.

High-resolution solar spectra,[23, 24, 25, 26, 27, 28] which have been taken for many years, represent the best source of information on atmospheric absorption lines. While these measurements have generally been made at high altitudes in order to minimize atmospheric effects and do not provide absolute data on transmission through a standard atmosphere, the measured lines at which attenuation occurs

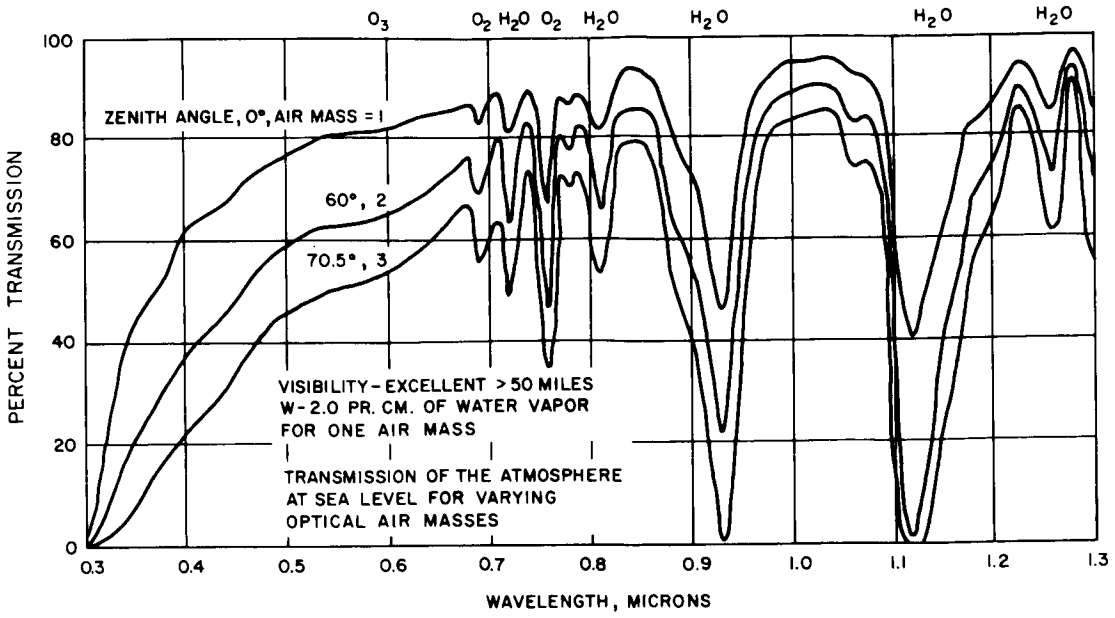


Figure 2-18. Transmission of the atmosphere at sea level for several zenith angles in the visible and near infrared.

are still strong and serve to identify those wavelengths which must be avoided in the design of a ground-based laser communication link. A detailed study of the absorption spectrum in the vicinity of a number of laser lines has been made[29, 30, 31] and indicates, for example, that the output of a ruby laser operated at room temperature lies between water vapor absorption lines situated about a half an angstrom unit above and below the 6943 Å operating point. While similar windows exist above and below this wavelength, the allowable operating temperature range for the laser rod to maintain the radiation within any one of these windows is of the order of 15°C. This emphasizes the importance of detailed spectral measurements of atmospheric absorption about each specific wavelength of interest.

Provided the laser wavelength does not coincide with an absorption line, attenuation in the atmosphere will be due to scattering effects. The attenuation at short wavelengths is due to molecular (Rayleigh) scattering of the radiation or, which the scattering

coefficient varies as $1/\lambda^4$, and together with absorption by ozone in the upper atmosphere accounts for the sharp cutoff of transmission in the ultraviolet as shown in Figure 2-18. Scattering from aerosol particles and droplets in the first few kilometers of the lower atmosphere also plays a major part in attenuation of electromagnetic radiation in the visible and near-infrared regions.[32, 33] For this type of (Mie) scattering (where particle dimensions are comparable with wavelength) the wavelength dependence of the scattering coefficient is a function of particle size and type, but for typical aerosol distributions encountered, experimental measurements[34] suggest that the dependence is about $1/\lambda$.

While little quantitative information is available on laser beam penetration of fog, clouds, or rain, the great increase in scattering of optical radiation associated with such weather conditions indicates that attenuation will be prohibitory. Some idea of the effects of fog and rain is given in Figure 2-19 based on measurement[35] of signal losses over a fixed laser link of 2.6 km. Note that the loss for rain ranges over 5 to 15 db versus 20 to greater than 60 db for fog. Attenuation in clouds[36] can be expected to be similar to that for fog.

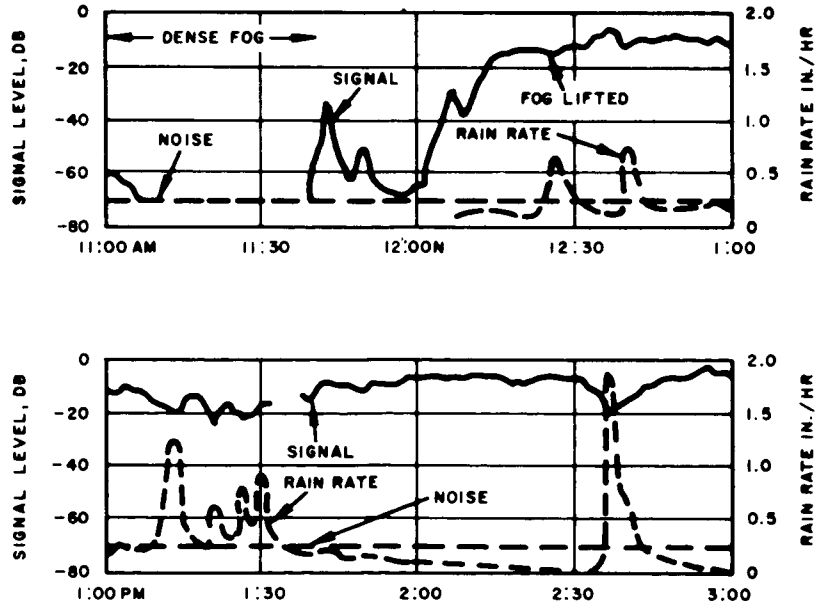


Figure 2-19. Fog-rain attenuation. $\lambda = 0.63\mu$, 2.6 km path length.

2.5 ATMOSPHERIC DISTORTION

The subject of electromagnetic propagation through turbulent refractive irregularities has received considerable attention as a result of investigations into the phenomena of beyond-the-horizon scatter propagation in the RF region and, of longer standing, in connection with astronomical seeing limitations in the optical region. More recently, the subject has become of concern in the development of deep space communication links where the turbulent characteristics of the atmosphere are beginning to impose serious limitations on tracking and communication capabilities for the large apertures and high frequencies considered. However, because of the complex nature of atmospheric perturbations and of scattering theory and because of the difficulties involved in carrying out meaningful experiments, quantitative predictions of phase and amplitude fluctuations over the propagating wavefront cannot yet be made reliably with reference to the pertinent meteorological conditions.

2.5.1 Distortion At Radio Frequencies

For RF propagation through the troposphere variations in the index of refraction are principally due to water vapor irregularities. These irregularities can be attributed to the turbulent movements of the random atmospheric wind velocity fields. In the stratosphere, which contains a negligible amount of water vapor as compared to the troposphere, variations in the refractive index are less and can be assumed to be primarily a function of thermodynamic variations in pressure and temperature. Models have been developed, describing the atmospheric structure in statistical terms, as a basis for the application of scattering theories.[37-41] Some correlation has been achieved between these theories and a number of experimental measurements[42-47] taken over various path lengths under different and incompletely defined meteorological conditions. A more definitive experiment is required to provide good correlation between meteorological conditions and measured effects on propagation. Such a program

is presently underway,[48] conducted by the Stanford Research Institute, and is designed specifically to determine propagation limitations on multiple aperture antennas for deep space communications.

In the absence of definitive data, an indication of expected deviations of an incoming wavefront over an antenna aperture can be gained from the phase measurements obtained in the NBS Maui experiment.[45] Here phase deviations in transmission over a 15 mile path, which dropped from 10,000 feet to 100 feet (corresponding to a zenith angle of about 83^{o}) were measured by two receivers for several baseline lengths up to 4800 feet. The measurements were made at a frequency of 9,414 MHz, but the deviations in terms of path length can be expected to a first approximation to be independent of frequency. Since the measurements at the various antenna locations were made during different recording periods and therefore under different meteorological conditions, the sample points are not strictly correlated over the baseline range. The trend of these samples, however, is remarkably consistent and is shown in Figure 2-20 where phase deviations have been converted to linear deviations in the wavefront.

2.5.2 Distortion At Optical Frequencies

Whereas in the RF region of the spectrum, humidity plays a predominant part in atmospheric propagation phenomena, in the optical region the index of refraction is determined essentially by the density of the air[49]. Wavefront distortions result principally then from local variations in density due to atmospheric temperature fluctuations.

The problem of phase and amplitude fluctuations in the signal wavefront is conceptually simpler at optical frequencies where propagating wavelengths are always very much smaller than the diffracting atmospheric inhomogeneities. At the same time, experimental measurements are made somewhat easier by the smaller correlation distances involved. As a result the theory of optical

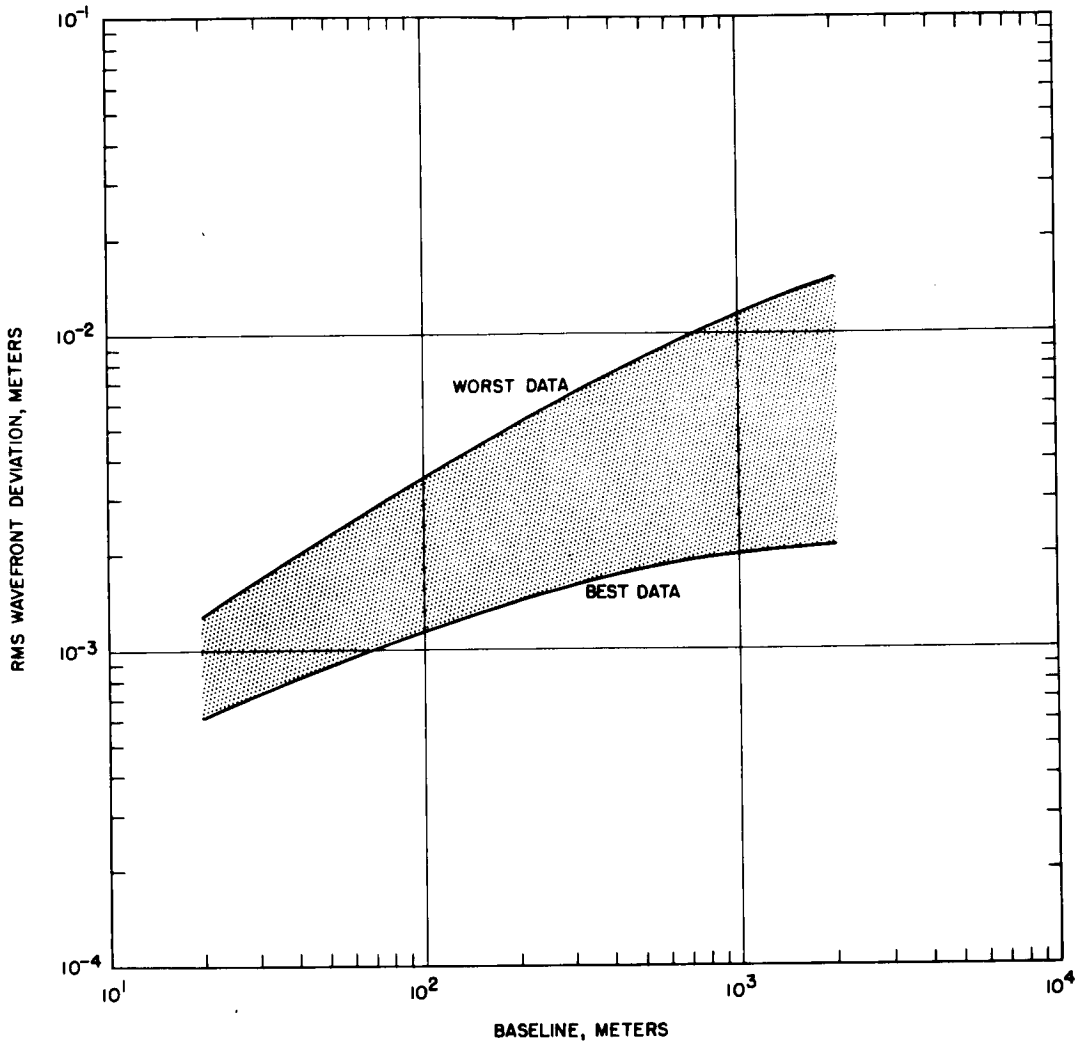


Figure 2-20. Atmospheric distortion of rf wavefront based on NBS Maui data.

propagation through the atmosphere[50] is well advanced and shows reasonable correlation with astronominal experience and with recent experiments using laser beams.[51, 52]

The effects of these atmospheric fluctuations on laser systems have been discussed[53, 54] and a major study program[55] has been undertaken which considers the specific problems associated with laser space communications. As a part of this study the lateral phase correlation length has been calculated[56] for typical daytime and

nighttime atmospheric conditions as shown in Figure 2-21.* The effective (elemental) aperture of a coherent receiver operating within the atmosphere is restricted to a diameter comparable with this correlation length as indicated in Figure 2-22. (A similar argument holds for an optical transmitter where the effective antenna gain is constrained by atmospheric distortions of the transmitted wavefront.) If larger effective aperture areas are to be achieved for coherent ground-based systems, self-phasing techniques will be required to provide correlation among a number of elements.

Amplitude fluctuations are also present in the received wavefront as a result of atmospheric disturbances. Although these do not represent a fundamental restriction on system performance, physical apertures considerably greater than the effective aperture limit may be necessary to average over several amplitude correlation lengths. In the visible region, for example, where the effective aperture diameter for coherent detection is only a few centimeters, the amplitude correlation length is typically of the order of 10 centimeters and a 30 centimeter aperture is required to reduce the rms deviations in received signal power to about 10 percent.[50] However, since the amplitude correlation length is proportional to $\lambda^{1/2}$ while the phase correlation length is proportional to $\lambda^{6/5}$, the two become comparable in the far infrared. At the same time the magnitude of the amplitude fluctuations across the wavefront decreases with increasing wavelength so that physical apertures need not be appreciably in excess of effective apertures for good performance at 10.6 microns.

For noncoherent detection effective aperture area is not restricted by phase deviations, but, as for coherent detection, apertures large with respect to amplitude correlation lengths are required to minimize signal fluctuations due to amplitude scintillations.

---

*Work[57] in progress at Hughes Research Laboratories indicates that the phase coherence length at 10.6 microns may be considerably greater than that predicted in Figure 2-21. Such a result would mean that the restriction on receiver diameter imposed by atmospheric turbulence would be much less severe at 10.6 microns than assumed here.

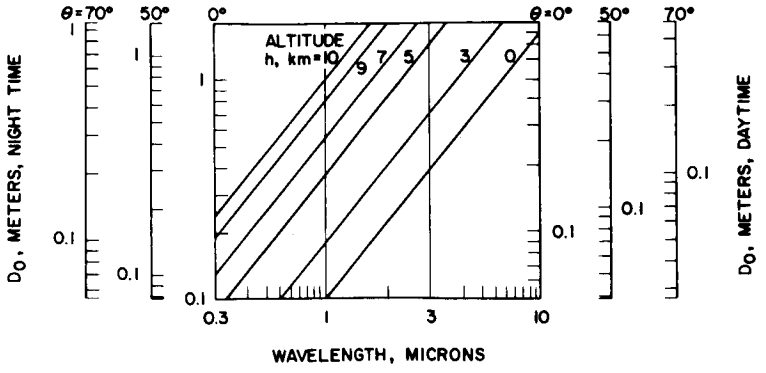


Figure 2-21. Dependence of correlation length, D_O, on zenith angle, θ, altitude, h, and wavelength, λ.

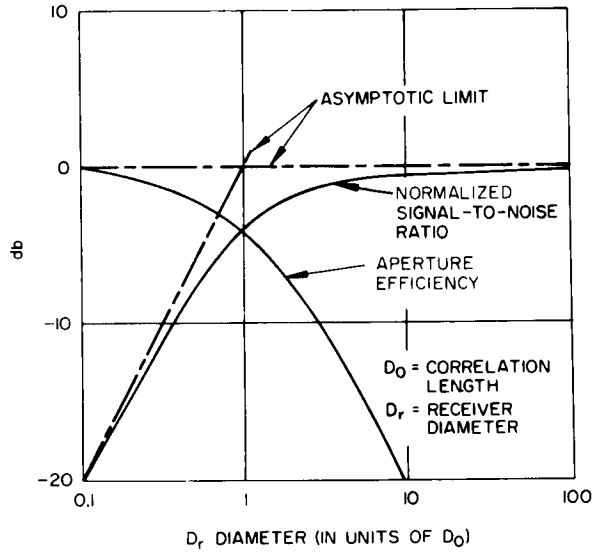


Figure 2-22. Signal loss for coherent detection caused by atmospheric turbulence.

Atmospheric turbulence is strongly influenced by characteristics of the surrounding terrain. Thus careful choice of site can reduce atmospheric effects by a factor of two or three. A similar advantage can be gained by elevation of the receiver above the turbulent layer that blankets the ground. The appropriate height[58] is a function of wind velocity and type of terrain; it is generally greater than 100 feet but may be as low as 20 feet under low-velocity, flat-terrain conditions.

2.6 REFERENCES

1. B. M. Oliver, Thermal and Quantum Noise, Proc. IEEE 53, 5, pp 436-454 (1965).

2. R. C. Hansen and R. G. Stephenson, Communications at Megamile Ranges, Micro. Jour. 4, December 1961.

3. R. W. Sanders, Communication Efficiency Comparison of Several Communication Systems, Proc. IRE 48, April 1960.

4. A. J. Viterbi, On Coded Phase-Coherent Communications, IRE Trans. SET 7, March 1961.

5. R. C. Hansen and R. G. Stephenson, Communications at Megamile Ranges, Micro. Jour. 5, January 1962.

6. L. S. Stokes and K. L. Brinkman, Reference Data for Advanced Space Communication and Tracking Systems, Report P66-135, Contract No. NAS 5-9637, Hughes Aircraft Company ASG, Culver City, California, 1966.

7. H. C. Ko, The Distribution of Cosmic Radio Background Radiation, Proc. IRE 46, 1, pp 208-215 (1958).

8. R. G. Stephenson, External Noise, Space Communications, Edited by A. V. Balakrishnan, McGraw-Hill Book Company, Inc.

9. R. G. Conway, K. I. Kellerman, and R. F. Long, The Radio Frequency Spectra of Discrete Radio Sources, Monthly Notices, Royal Astron. Soc. 125, p 268-269 (1963).

10. A. G. Smith, Extraterrestrial Noise as a Factor in Space Communications, Proc. IRE, 48, 4, pp 594 (1960).

11. L.V. Blake, Tropospheric Absorption and Noise Temperature for a Standard Atmosphere, PT-GAP Int'l Sym. NBS, Boulder, Colo. 1963.

12. C. H. Mayer, Thermal Radio Radiation from the Moon and Planets, IEEE Trans. AP12, 7, pp 902-913 (1964).

13. J. E. Gibson, Lunar Thermal Radiation at 35 Kmc, Proc. IRE 46, 1, pp 280-286 (1958).

14. G. H. Millman, Atmospheric Effects on VHF and UHF Propagation, Proc. IRE 48, 8, pp 1492-1501 (1958).

15. R. C. Ramsey, Spectral Irradiance from Stars and Planets, above the Atmosphere, from 0.1 to 100.0 Microns, Appl. Optics 1, 4 (1962).

16. H. H. Malitson, The Solar Energy Spectrum, Sky and Telescope, pp 162-165, March 1965.

17. I. L. Goldberg, Radiation from Planet Earth, U. S. Army Signal Research and Development Laboratory, Ft. Monmouth, N. J., September 1961.

18. S. K. Mitra, The Upper Atmosphere, Asiatic Society, Calcutta, India, 1962.

19. F. E. Roach, Manual of Photometric Observations of the Airglow During the IGY, National Bureau of Standards Report No. 5006.

20. W. T. Hunt, Survey of Attenuation by the Earth's Atmosphere at Millimeter RadioWavelengths, WADD TN 60-232, ASTIA No. 252-126, November, 1960.

21. J. W. Meyer, Radar Astronomy at Millimeter and Submillimeter Wavelengths, Proc. IEEE 54, 4, pp 484-492 (1966).

22. R. M. Chapman and R. Carpenter, Effect of Night Sky Backgrounds on Optical Measurements, Tech. Rpt. 61-23-A, Geophysics Corp. of America, May 1961.

23. M. Minnaert, G. F. W. Mulders, and J. Houtgart, Photometric Atlas of the Solar Spectrum from λ3612 to λ8771, Schnabel, Kampert and Helm, Amsterdam, The Netherlands, 1940.

24. H. D. Babcock and C. E. Moore, The Solar Spectrum, 6600 to 13495, Carnegie Institution of Washington, Publication 579, 1947.

25. M. Migeotte, The Solar Spectrum Observed at the Junglraujoch (Switzerland) 7500 to 9070 Angstrams, Tech. Status Rpt. No. 16, Contract AF 61(514) -962, July-Sept. 1960.

26. M. Migeotte, Annex to Technical Status Report Number 18, Contract AF 61(514)-962, Jan. - March 1961.

27. O. C. Mohler, A. K. Pierce, R. R. McMath, and L. Goldberg, Photometric Atlas of the Near Infrared Solar Spectrum λ8465 to λ25242, University of Michigan Pres, Ann Arbor, Mich., 1950.

28. M. Migeotte, L. Neven, and J. Swensson, The Solar Spectrum from 2.8 to 23.7 Microns, Part I, Photometric Atlas, Tech. Final Rpt., Phase A, Part I, Contract AF 61(514)-432, 1957.

29. R. K. Long, Absorption of Laser Radiation in the Atmosphere, Rept. No. 1579-3, The Ohio State University Research Foundation, Contract AF 33(657)-10824, 196

30. R. K. Long, Atmospheric Attenuation of Ruby Lasers, Proc. IEEE 51, 5, pp 859-860 (1963).

31. R. K. Long and C. H. Boehnker, Measured Atmospheric Absorption at Ruby Optical Maser Wavelengths, Report 1641-10, the Ohio State University Research Foundation, Contract No. AF 33(657)-11195, 1965.

32. W. E. K. Middleton, Vision Through the Atmosphere, Univ. of Toronto Press, 1958.

33. L. Elterman, A Model of a Clear Standard Atmosphere for Attenuation in the Visible Region and Infrared Windows, Research Report, Optical Physics Laboratory Project 7670, AFCRL, 1963

34. G. L. Knestrick, T. H. Corden, and J. A. Curcio, Atmospheric Scattering Coefficients in the Visible and Infrared Regions, J. Opt. Soc. Am. 52, 9, pp 1010-1016 (1962).

35. D. C. Hogg, Effect of the Troposphere on the Propagation of Coherent Optical Waves, Proc. PTGAP IEEE International Symposium, pp 102-108, 1965.

36. D. Deirmendjian, Scattering and Polarization Properties of Water Clouds and Hazes in the Visible and Infrared, Appl. Optics 3, 2, pp 187-196 (1964).

37. H. S. Booker and W. E. Gordon, A Theory of Radio Scattering in the Troposphere, Proc. IRE 38, pp 401-412 (1950).

38. R. B. Muchmore and A. D. Wheelon, Line-of-Sight Propagation Phenomena - I. Ray Treatment, Proc. IRE 43, pp 1437-1449 (1955).

39. A. D. Wheelon and R. B. Muchmore, Line-of-Sight Propagation Phenomena-II. Scattered Components, Proc. IRE 43, pp 1450-1458 (1955).

40. A. D. Wheelon, Near Field Corrections to Line of Sight Propagation, Proc. IRE 43, pp 1459-1466 (1955).

41. A. D. Wheelon, Radio-Wave Scattering by Tropospheric Irregularities, J. of Res. N. B. S. 63 D, 2, pp 205-233 (1959).

42. J. W. Herbstreit and M. C. Thompson, Measurements of the Phase of Radio Waves Received over Transmission Paths with Electrical Lengths Varying as a Result of Atmospheric Turbulence, Proc. IRE 43, pp 1391-1401 (1955).

43. A. P. Deam and B. M. Fannin, Phase Difference Variations in 9,350-Megacycle Radio Signals Arriving at Spaced Antennas, Proc. IRE 43, pp 1402-1404 (1955).

44. M. C. Thompson and H. B. Janes, Measurements of Phase Stability over a Low-Level Tropospheric Path, J. of Res. N. B. S. 63 D, 1, pp 45-51 (1959).

45. K. A. Norton, et al., An Experimental Study of Phase Variations in Line-of-Sight Microwave Transmissions, N. B. S. Monograph 33, November 1961.

46. M. T. Decker and F. O. Guirand, Tropospheric Radio Propagation on an Oversea Path in the Gulf of Mexico, N. B. S. Report 8429, May 1964.

47. P. L. Smith, Scattering of Microwaves by Cloud Droplets, Proc. World Conference on Radio Meteorology, Boulder, Colorado, pp 202-207, 1964.

48. J. E. Adler, et al., The Design of an Experiment to Determine the Limitations Imposed on a Multiple Aperture Antenna System by Propagation Phenomena, Contract No. NAS 5-3974, SRI Project 5067, Stanford Research Institute, Menlo Part, California.

49. H. R. Reed and C. M. Russell, Ultra High Frequency Propagation, John Wiley & Sons, Inc., New York, 1953.

50. V. I. Tatarski, Wave Propagation in a Turbulent Medium, Translation from Russian, McGraw-Hill, 1961.

51. A. L. Buck, Laser Propagation in the Atmosphere, Proc. Conference on Atmospheric Limitations to Optical Propagation, pp 27-38, Boulder, Colorado, 1965.

52. I. Goldstein, A. Chabot, and P. A. Miles, Heterodyne Measurements of Light Propagation through Atmospheric Turbulence, Proc. Conference on Atmospheric Limitations to Optical Propagation, pp 273-292, Boulder, Colorado, 1965.

53. J. I. Davis, Consideration of Atmospheric Turbulence in Laser Systems Design, Appl. Optics 5, 1, pp 139-147 (1966).

54. H. Hodara, Laser Wave Propagation through the Atmosphere, Proc. IEEE 54, 3, pp 3 8-375 (1966).

55. Laser Space Communications Systems (LACE)Study, North American Aviation, Inc., SID, NASA Contract NAS w, 977.

56. D. L. Fried, The Effect of Wave Front Distortion on the Performance of an Ideal Optical Heterodyne Receiver and an Ideal Camera, Proc. Conference on Atmospheric Limitations to Optical Propagation, pp 192-241, Boulder, Colorado, 1965.

57. W. P. Brown, Jr., to be published.

58. E. K. Webb, Daytime Thermal Fluctuations in the Lower Atmosphere Appl. Optics 3, 12, p 1329 (1964).

3.0 SYSTEM PERFORMANCE

The potential telecommunication performance for a deep space communication link (from a space vehicle to earth) is considered in this section as a function of frequency. Consideration is given to both the basic and practical limitations which apply.

First, as a measure of telecommunication system performance, the product of signal quality and quantity is derived from the one-way transmission equation, and the explicit frequency-dependent terms are separated from the system parameters. Next, link performance is investigated as a function of the explicit frequency dependency and as it is influenced by the implicit restrictive frequency dependencies imposed on the "constant" parameters of the transmission link equation. Included are the losses due to atmospheric absorption and the influence of atmospheric distortion on receiver aperture size. Total link performance is then given as a product of the various frequency dependent parameters. The optimum operating frequencies indicated as a result of this analysis will depend on the relative values taken as parametric restrictions. Finally, consideration is given to operational system requirements and their effect on the overall system implementation. This consideration includes the relative role of earth receiving stations versus satellite receiving stations.

3.1 PERFORMANCE CRITERIA

The performance of a telecommunication link may be measured against many criteria. Chief among these criteria are amount of data, trustworthiness, and economy in transmission of the data. In this study, primary emphasis is placed on achieving an improvement in the amount of data while maintaining its quality. Economy of data transmission remains important but is placed in a secondary role at this time, for in general, realistic cost evaluations must be related to specific candidate systems as they develop from basic systems studies and research. The product of transmitted data quality and quantity (the information rate parameter) is then taken as a performance criterion,

the one way transmission equation is rewritten in terms of this product. Here the signal-to-noise ratio, S/N, represents signal quality; the bandwidth, B, represents signal quantity:

$$\mathcal{Q}_o = \left(\frac{S}{N}\right)(B) = \frac{G_T C_R P_T \lambda^2}{(4\pi)^2 R^2 L\, hf \left[1 + (e^{hf/kT} - 1)^{-1}\right]} \qquad (1)$$

Insertion of the functional frequency dependencies of the transmitter and receiver antenna gain parameters of this equation ($G = 4\pi A/\lambda^2$) gives:

$$\mathcal{Q}_o = \left(\frac{S}{N}\right)(B) = \frac{A_T A_R P_T}{R^2 hc^2 L_o L_f} M \qquad (2)$$

where L_o and L_f represent the fixed and frequency dependent losses (greater than unity), respectively, and M^1 collects the functional frequency dependencies,

$$M = \frac{f}{1 + (e^{hf/kT} - 1)^{-1}} \qquad (3)$$

Equation 2 separates the explicit functional frequency dependencies from the physical parameters of the transmission link equation. However, some of these physical parameters are indirectly dependent on frequency in terms of practical system considerations. Consider, for example, the frequency dependency placed on a spaceborne transmitting antenna due to achievable pointing accuracy. Such a limit sets a minimum allowable beamwidth (or maximum gain), thereby requiring the effective diameter of the transmitting antenna to decrease inversely as frequency increases. This and other such restrictions must be included when applying the transmission equation to the process of optimum frequency selection. The explicit functional frequency dependency given by M is

discussed below as are the implicit frequency dependencies due to the various practical limitations imposed on the transmitting antenna area, A_T; the receiving antenna area, A_R; the transmitted power, P_T; and losses due to atmospheric transmission represented by L_f.

3.1.1 Functional Frequency Dependency, M

Equation 3 compares ultimate system performance for communication systems operating at different frequencies but with constant values of antenna area for both transmit and receive, constant power transmitted, and constant losses in transmission. This equation is plotted in Figure 3-1 and indicates that with the above physical parameters fixed, the performance increases linearly with frequency where quantum noise is dominant and with the square of frequency where thermal noise is dominant. As indicated in this figure, there is an ultimate system improvement possible of six orders of magnitude in going from present operating frequencies in the microwave band to the visible light spectrum. While the actual achievement of this million-fold improvement in system performance is inhibited by the practical parametric restrictions to be discussed, still the frequency-dependent parameter M does represent an ultimate performance limit, with respect to fixed physical parameters which could be approached upon the investment of suitable effort. It sets a goal, the pursuit of which will require the development of new concepts and techniques.

3.1.2 Transmitting Antenna Area

The effective area of a spacecraft transmitting antenna may be limited by one of three practical considerations:

- weight and size limitations imposed by the vehicle,
- antenna fabrication tolerances, and
- transmitting beamwidth demanded by the limitation on pointing accuracy.

Although there is some interrelation among these factors (for example, an extensible antenna may be used for increased diameter at the

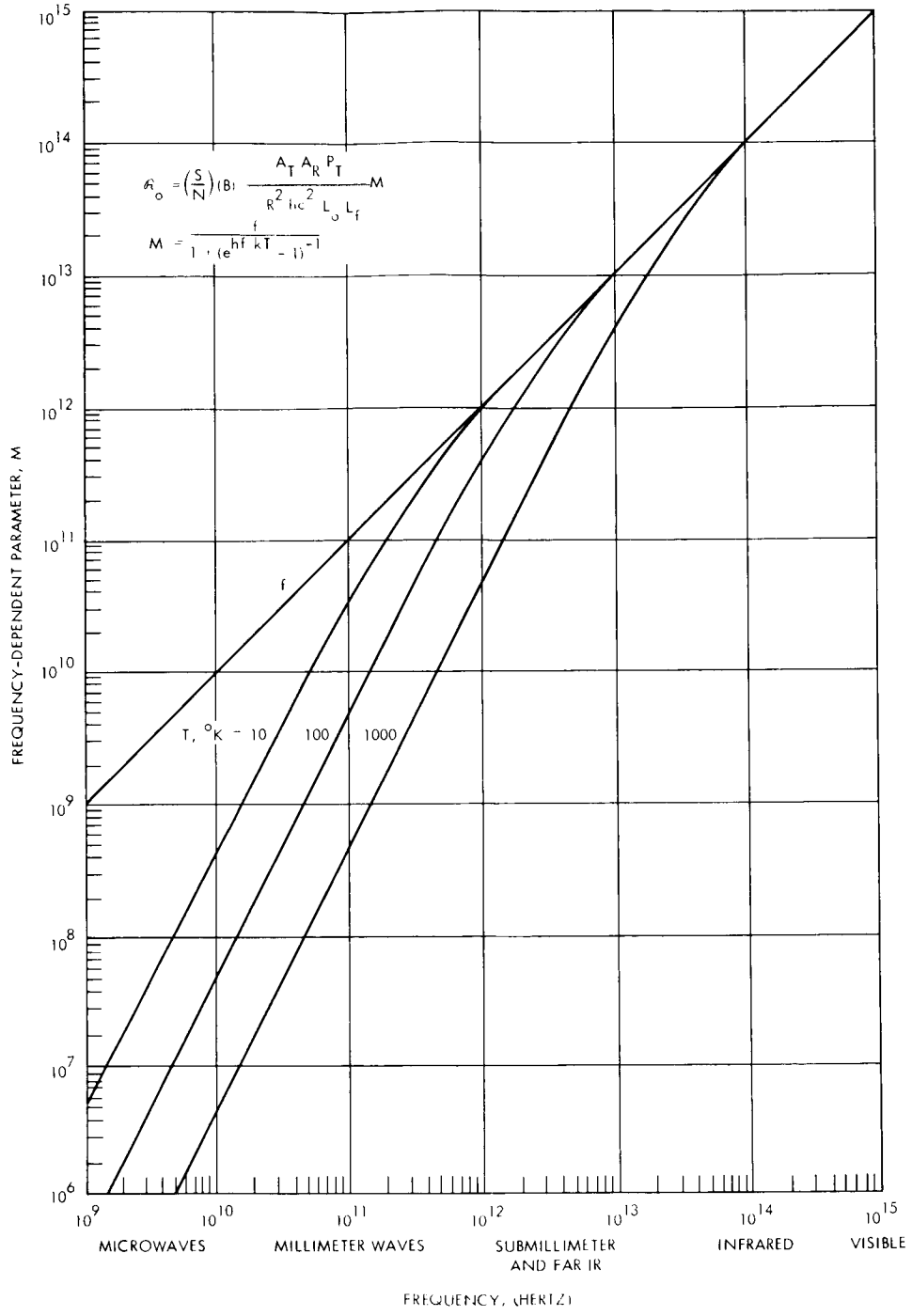


Figure 3-1. Functional frequency dependency of system performance (physical parameters fixed).

expense of dimensional tolerances) one will generally dominate within a certain frequency range.

In Figure 3-2 transmitting antenna area is plotted against frequency for different values of the three restrictive conditions. Horizontal lines indicating restrictions on effective diameter could be due to a variety of constraints such as vehicle dimensions or weight allocations. Fabrication tolerances set a limit on achievable gain and hence on effective aperture area. Tolerances are generally expressed in normalized form as the ratio of the rms dimensional deviation in effective path length to the diameter, σ/D. Since a given ratio implies a particular gain limit[2], effective antenna area will be inversely proportional to the square of the frequency for a given σ/D value. Similarly, a given pointing accuracy requires a certain beamwidth and hence represents a gain limit. Thus, the same family of curves can represent the restriction on transmitting antenna area due to either fabrication tolerance or pointing accuracy limitations. The curves, evaluated for an approximate 3 db loss, result in lines of constant negative slope on the log-log plot of Figure 3-2 and are labeled in terms of σ/D, beamwidth, and effective gain (3 db loss).

With reference then to Figure 3-2, the frequency dependency of the transmitting antenna area may be determined based on an appropriate set of assumptions as to vehicle constraints, practical fabrication tolerances, and achievable pointing accuracy.

3.1.3 Receiving Antenna Area

The same type of restrictions apply to the receiving antenna area as to the transmitting antenna. Thus, the useful area of a single element may be limited by allowable size, achievable tolerances, or (for coherent reception) achievable pointing accuracy. For a satellite receiving station, limiting values may not differ too much from those for a deep space vehicle. On the other hand, for a ground station, overall aperture diameter will in general not be restricted by installation requirements, but a single aperture element will be subject to a limitation due to atmospheric distortion of the signal wavefront over the aperture.

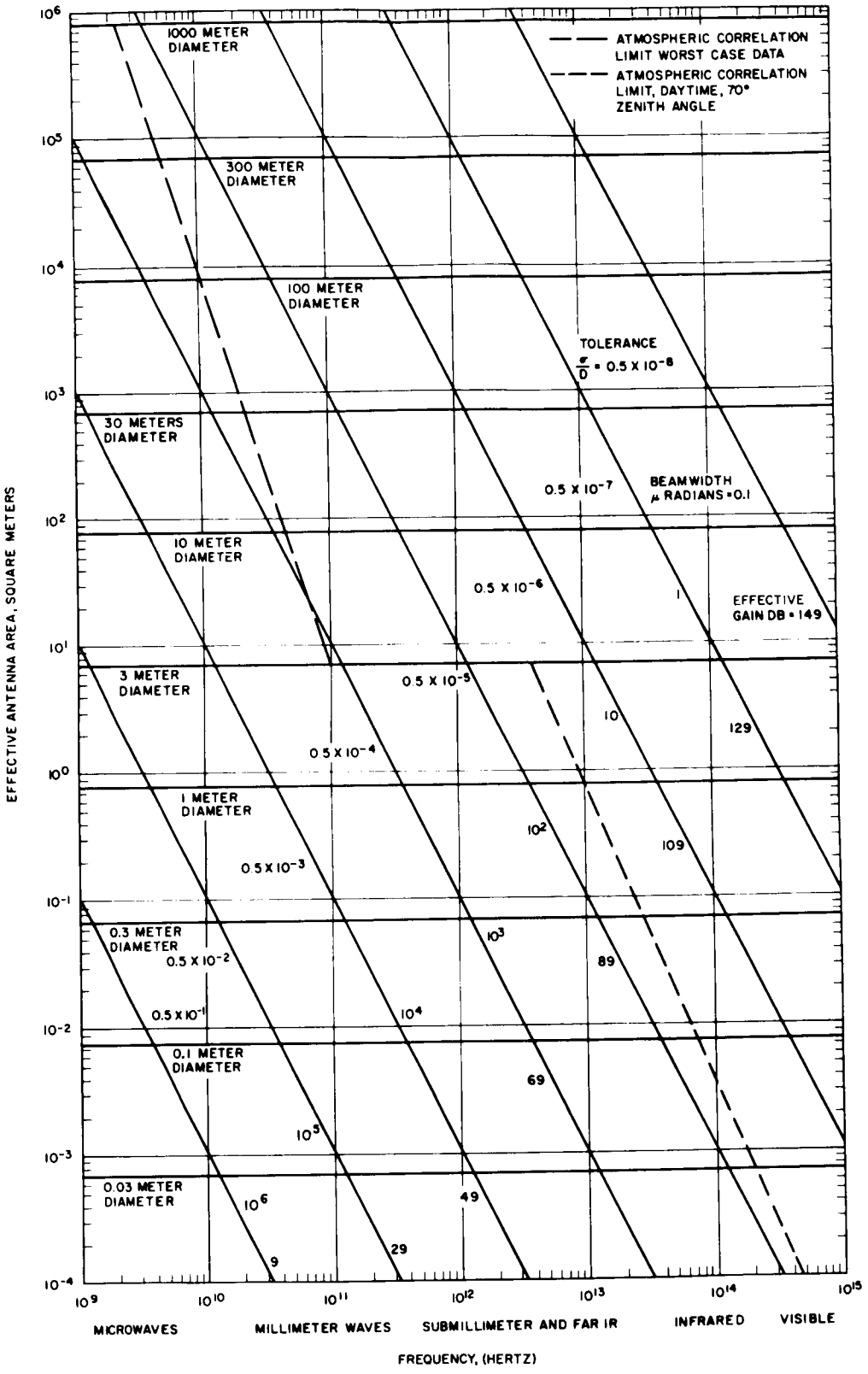


Figure 3-2. Parameter restrictions on effective antenna area.

As discussed in Section 2.5, little quantitative information is available on such atmospheric anomalies as wavefront corrugation and tilt, or unequal illumination of the aperture. However, an indication of the restriction imposed at radio frequencies on antenna diameter can be gained based on data available from the NBS experiment[3] at Maui, Hawaii (1956). The assumption is made that the wavefront distortions, corresponding to the phase deviations experimentally measured as a function of baseline (see Figure 2-20), will have an effect similar to that caused by dimensional deviations of the antenna surface due to fabrication tolerances. It is further assumed the linear dimensions of the distortion are to a first approximation independent of frequency[4]. A curve based on the worst data from the Maui experiment is plotted with the other restrictions in Figure 3-2.

In the optical region of the spectrum an atmospheric correlation length has been defined equal to the maximum effective aperture diameter for coherent detection[5]. This correlation length is assumed to vary as $\lambda^{6/5}$. The aperture area corresponding to a 3 db loss due to signal decorrelation for typical daytime conditions of atmospheric turbulence is plotted in Figure 3-2. This restriction applies only to coherent reception.

3.1.4 Transmitted Power

Frequency dependent restrictions on transmitter power are related to the method of power generation. At microwave frequencies traveling-wave tubes have been found most satisfactory in providing high powers, and this capability extends to the millimeter wavelengths. In general, however, maximum power generation capability in TWT amplifiers is inversely proportional to frequency.

At optical frequencies the principal limitation on power generated is set by heat dissipation in the laser cavity as a result of low conversion efficiencies. There is no significant systematic dependency on frequency, performance being related to the different specific laser mechanisms.

In either portion of the spectrum an increase in power can always be achieved by paralleling units although in the optical region phase-locking of the various elements will be more difficult. Hence the effective limitation is set by weight, size, and power supply restrictions.

3.1.5 Atmospheric Losses

Transmission losses occur in the atmosphere due to absorption of electromagnetic radiation by the constituent gases (especially water vapor) and, in the optical region, due to aerosol and molecular scattering. Absorption occurs in bands in the spectrum rather than as a systematic variation; this spectral absorption is indicated in Figures 2-14 through 2-18, in Section 2.4. Scattering at low altitudes from aerosols generally accounts for the predominant loss in the visible region, where the transmission is roughly proportional to wavelength as indicated in Figure 2-18.

Rain and fog begin to have a serious effect on transmission in the millimeter region of the RF spectrum, with the effect due to fog increasing at optical frequencies as indicated in Figures 2-14 and 2-19. It is evident that an optical communication link would be effectively blacked out by cloud cover. Table 3-1[6] shows the percentage of time that various sites experience obscuration by clouds and indicates the seriousness of this problem.

3.1.6 System Noise

Essentially system noise has no explicit frequency dependency other than that already included in the parameter M. It depends in practice, however, on external noise sources and on achievable receiver noise temperatures in the RF region and detector quantum and heterodyne efficiencies in the optical region. Presently achievable receiver performances in the RF and projected performance in both the RF and optical region are indicated in the Figure 3-3.[1] A receiver noise figure was not explicitly included in the one-way transmission equation, but has the effect of

Station Coordinates (Long.)	Lat.		Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
New Mexico		1962	36	25	27	15	5	12	49	32	48	19	26	22
253°27'(E)	+32°25'	1963	34	27	21	29	30	13	48	57	28	16	27	17
South Africa		1962	42	29	31	26	4	6	1	5	8	29	53	37
028°15'(E)	-25°58'	1963	50	23	27	28	15	21	20	2	7	30	55	49
Australia		1962	32	21	21	9	34	12	22	22	27	38	15	18
136°46'(E)	-31°06'	1963	16	22	22	22	51	35	47	21	18	17	18	22
Spain		1962	52	32	81	43	33	15	21	15	34	47	40	43
353°48'(E)	+36°28'	1963	72	49	39	39	35	32	4	19	22	19	60	45
Peru		1962	81	71	56	38	10	2	5	6	35	22	42	55
288°30'(E)	-16°28'	1963	88	95	63	56	23	2	6	15	21	26	13	52
Iran		1962	28	44	33	54	10	1	18	16	03	nil	14	38
052°31'(E)	+29°38'	1963	17	40	30	47	30	7	15	23	1	12	35	22
Curacao		1962	52	43	56	53	74	66	46	38	60	39	45	50
291°10'(E)	+12°05'	1963	70	59	51	68	70	40	64	55	63	50	65	62
Florida		1962	55	40	62	55	36	61	41	55	52	45	55	43
279°53'(E)	+27°01'	1963	57	58	47	28	50	42	44	29	49	32	49	43
Argentina		1962	53	52	38	72	51	27	39	32	22	29	35	43
294°54'(E)	-31°57'	1963	*	35	40	27	39	33	36	32	57	46	46	30
Hawaii		1963	23	41	56	39	33	5	38	18	20	16	26	33
203°45'(E)	+20°43'	1963	59	17	70	78	52	28	29	23	39	39	32	30

*No photography attempted for a 3 week period when the mirror was removed for realuminizing.

Table 3-1. Baker Nunn sites showing percent of viewing time lost due to clouds.

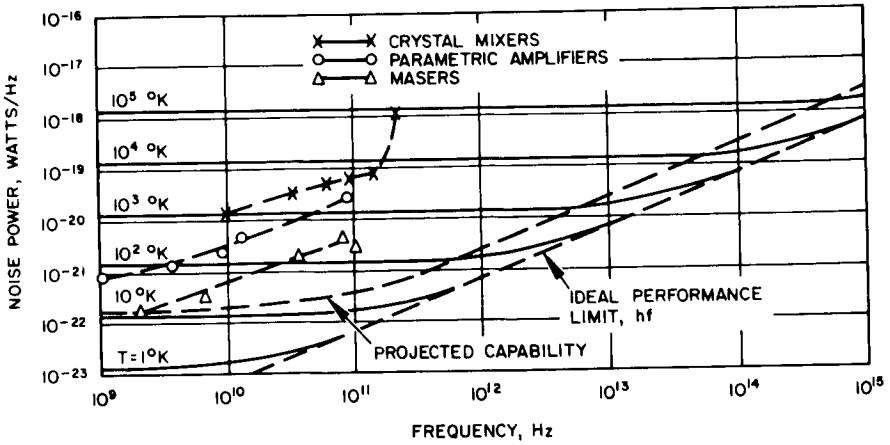


Figure 3-3. Heterodyne receiver noise performance.

of reducing the value of the parameter M. The noise figure and external noise (see Section 2) can be accounted for in the RF region by taking the noise-equivalent rather than actual input temperature to give the effective value of M. In the optical region M must be reduced by the quantum efficiency of the detector as well as by the receiver noise figure when considering the performance to be expected of a real system. The effective value of M has been taken, therefore, as 6 db below the value indicated in Figure 3-1, the loss being equally divided between the detector quantum efficiency and the receiver noise figure.

3.1.7 Fixed Losses

In addition to frequency and weather dependent transmission losses and to detection losses accounted for by the effective value taken for the parameter M, there are several practical losses which apply generally to all systems. The limitations on effective transmitting and receiving apertures have been given for conditions to which

a 3 db decrease in effective gain is expected; thus a 6 db loss must be allocated to these components. A further 1 db is then taken to cover miscellaneous transmission losses in the equipment, so that a total of 7 db is assumed for fixed system losses.

3.2 SAMPLE PERFORMANCE CALCULATION

System performance versus frequency may now be calculated based on the one-way transmission equation and on various limiting values of the parameters as discussed above. A calculation of performance at a range of one astronomical unit is made taking representative parametric restrictions as tabulated in Table 3-2. In general, the technological restrictions represent the state of the art projected to the 1975-80 period. The curves are normalized, however, to a transmitted power of one watt, the assumption being made that if the requirement exists, roughly the same transmitter efficiency can be obtained in any region of the spectrum. The curves are also based on a single element receiver aperture. Practical restrictions imposed by the atmosphere define three cases of interest:

- no atmospheric losses (corresponding to a satellite receiving station),
- atmospheric losses due to clear weather conditions, and
- atmospheric losses due to poor weather conditions.

The specific parameter values for several frequencies particularly those at which significant changes occur in the chosen parametric restrictions are given in Table 3-3. The resultant product of signal quality (S/N) and signal quantity (B) is also tabulated.

The product (S/N) (B), or information-rate parameter $\mathcal{R}_o$, is plotted in Figure 3-4 for the three cases considered with reference to the atmosphere. The sharp peaks and valleys result from the discontinuous nature of the parametric restrictions selected, and their exact positions are dependent on the particular values taken as representative of practical limits.

Frequency Dependent Parameter	No Atmospheric Losses (Satellite Receiving Station)	Atmospheric Losses, Clear Weather	Atmospheric Losses, Including Rain and Fog
Power transmitted, P_T	Power taken as one watt for all frequencies		
Transmitting antenna area, A_T, limits (see Figure 3-2)	10 meter maximum diameter effective to frequency of 3×10^9 cps; $\sigma/D = 0.64 \times 10^{-4}D$ to frequency of 3×10^{11} cps; 1 meter maximum diameter to 4×10^{13} cps; beamwidth of 10 μ radian to 10^{15} cps.		
Receiving antenna area, A_R, limits (see Figure 3-2)	Same as A_T limits above	100 meter diameter to 10^{10} cps; atmospheric correlation limit to 10^{11} cps; 3 meter diameter to 4×10^{12} cps; atmospheric correlation limit to 10^{15} cps.	
Atmospheric frequency dependent losses, L_f	None	Losses per Figures 2-15, 2-16, 2-17 and 2-18 (80° Zenith angle for RF, 70° Zenith angle, 32 km visibility for optical).	Losses per Figures 2-14 and 2-19 (80° Zenith angle for RF, 70° Zenith angle for optical).
System noise	Antenna temperature per Figure 2-2 (maximum galactic background); receiver noise temperature 1/2 paramp state of the art per Figure 3-3.	Antenna temperature per Figure 2-2 and Figure 2-7 (80° Zenith angle) to 5×10^{10} cps, 300°K to 2×10^{13} cps; receiver noise per Figure 3-3 (projected capability).	Antenna temperature per Figure 2-2 and Figure 2-7 (80° Zenith angle) to 3×10^9 cps, increasing to 300°K at 10^{10} cps, 300°K to 10^{15} cps; receiver noise per Figure 3-3 (projected capability).

Table 3-2. Sample frequency dependent parameter limitations.

Frequency, Hz	$1/R^2hc^2L_o$, watts^{-1} meters^{-4}	P_T, watts	A_T, meters^2	A_R, meters^2	$1/L_f$	M, Hz	(S/N) (B), Hz
10^9	1.49×10^{-7}	1	78	78	1.0	8.4×10^5	7.6×10^2
3×10^9	↓	↓	78	78	↓	1.6×10^6	1.5×10^3
10^{10}			23.4	23.4		7.6×10^7	6.2×10^3
2.2×10^{10}			10.6	10.6		1.7×10^8	2.8×10^3
3×10^{10}			7.8	7.8		2.3×10^8	2.1×10^3
10^{11}			2.3	2.3		7.6×10^8	6.2×10^2
3×10^{11}			0.78	0.78		2.3×10^9	2.1×10^2
3×10^{13}			0.78	0.78		7.5×10^{12}	6.8×10^5
3×10^{14}			1.3×10^{-2}	1.3×10^{-2}		7.5×10^{13}	1.9×10^3
10^{15}	1.49×10^{-7}	1	1.2×10^{-3}	1.2×10^{-3}	1.0	2.5×10^{14}	5.0×10

Table 3-3a. System parameter values, no atmosphere.

Frequency Hz	$1/R^2hc^2L_o$, watts^{-1} meters^{-4}	P_T, watts	A_T, meters^2	A_R, meters^2	$1/L_f$	M, Hz	(S/N) (B), Hz
10^9	1.49×10^{-7}	1	78	7.8×10^3	0.96	7.9×10^5	6.8×10^6
3×10^9			78	7.8×10^3	0.96	1.5×10^7	1.3×10^6
10^{10}			23.4	7.8×10^3	0.94	1.4×10^8	3.6×10^6
2.2×10^{10}			10.6	6.8×10^2	0.42	1.3×10^8	5.9×10^4
3×10^{10}			7.8	2.7×10^2	0.76	4.5×10^8	1.1×10^5
10^{11}			2.3	7.1	0.32	1.5×10^9	1.2×10^3
3×10^{11}			0.78	7.1	$\sim 10^{-4}$	1.2×10^{10}	$\sim 9.5 \times 10^{-1}$
3×10^{13}			0.78	7.5×10^{-2}	0.6	7.5×10^{12}	3.9×10^4
3×10^{14}			1.3×10^{-2}	2.9×10^{-4}	0.6	7.5×10^{13}	2.6×10
10^{15}	1.49×10^{-7}	1	1.2×10^{-3}	1.7×10^{-5}	—	2.5×10^{14}	—

Table 3-3b. System parameter values, clear weather.

Frequency, Hz	$1/R^2hc^2L_o$, watts^{-1} meters^{-4}	P_T, watts	A_T, meters^2	A_R, meters^2	$1/L_f$	M, Hz	(S/N) (B), Hz
10^9	1.49×10^{-7}	1	78	7.8×10^3	0.96	7.9×10^5	6.8×10^4
3×10^9	↓	↓	78	7.8×10^3	0.96	1.5×10^7	1.3×10^6
10^{10}			23.4	7.8×10^3	0.5	1.5×10^7	2.0×10^5
2.2×10^{10}			10.6	6.8×10^2	0.04	7.4×10^7	3.2×10^3
3×10^{10}			7.8	2.7×10^2	0.02	1.4×10^8	8.8×10^2
10^{11}			2.3	7.1	$\sim 10^{-5}$	1.4×10^9	$\sim 10^{-2}$
3×10^{11}			0.78	7.1	—	1.2×10^{10}	—
3×10^{13}			0.78	7.5×10^{-2}	—	7.5×10^{12}	—
3×10^{14}			1.3×10^{-2}	2.9×10^{-4}	—	7.5×10^{13}	—
10^{15}	1.49×10^{-7}	1	1.2×10^{-3}	1.7×10^{-5}	—	2.5×10^{14}	—

Table 3-3c. System parameter values, rain and fog.

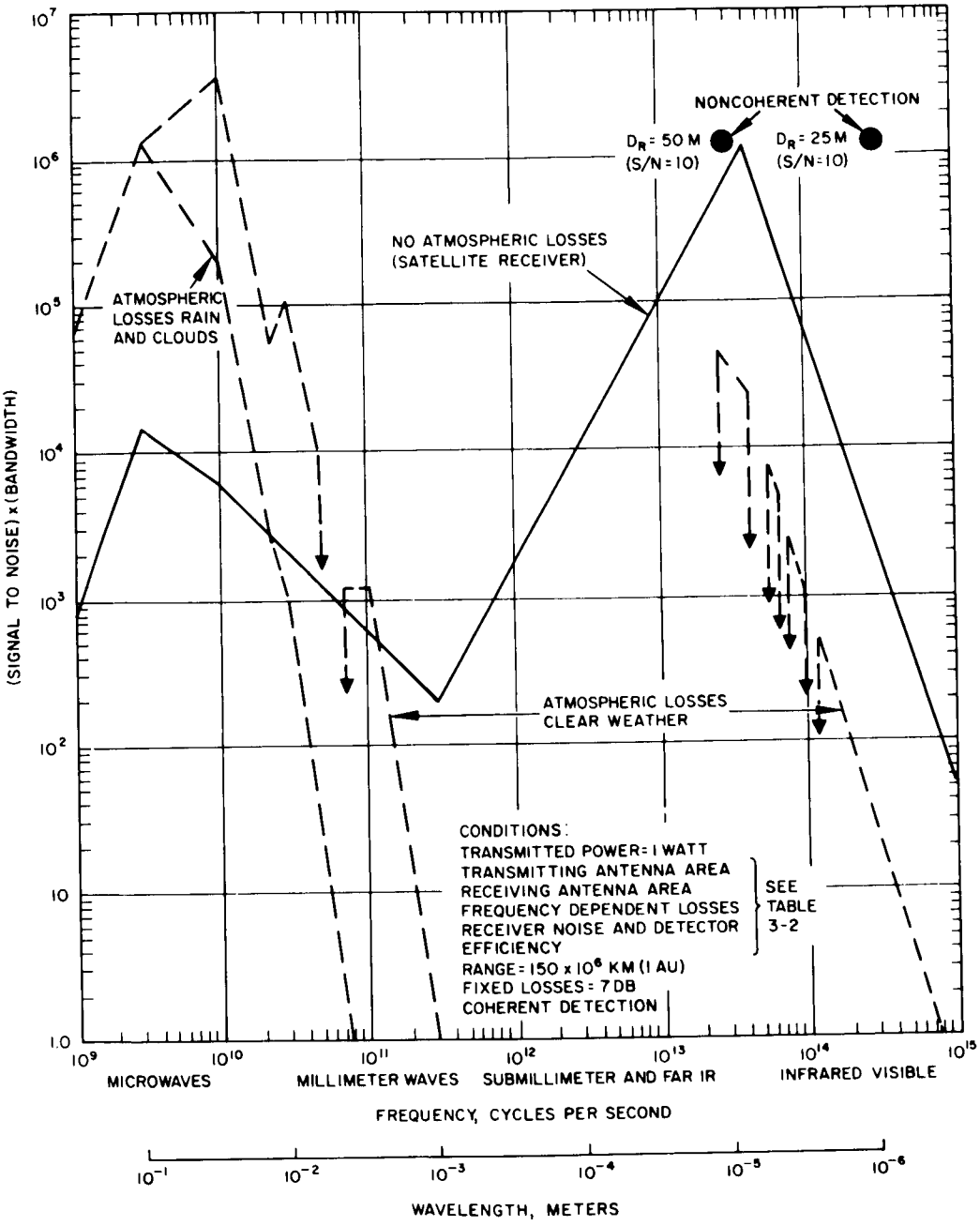


Figure 3-4. Communication system performance.

It should also be noted that the curves in the optical region represent performance limitations on coherent detection. For non-coherent detection, pointing accuracy and atmospheric distortions do not impose restrictions on receiver aperture area so that in general the fall off in performance with frequency in the optical region (for no atmosphere or for clear weather) will be as the inverse of frequency rather than the inverse cube. On the other hand, receiver noise levels may be very much higher. A comparison is made between coherent and noncoherent performance at two specific wavelengths as shown in the figure.

3.2.1 Satellite Receiver

The curve for a satellite receiving station shows two maxima, in the vicinity of 3×10^9 and 3×10^{13} Hz, corresponding to wavelengths of approximately 10 centimeters and 10 microns, respectively. These are directly related to the maximum frequencies for which constant aperture areas were taken for both the transmitter and receiver. The falloff in performance on the low-frequency side of the microwave peak results from a direct limitation on aperture dimension while the cutoff on the high-frequency side of this peak results both from a restriction on antenna size due to fabrication tolerances (down to dimensions where optical fabrication techniques apply) and from the higher receiver noise temperatures appropriate to an uncooled satellite receiver in the millimeter-wave region of the spectrum. The cutoff on the high-frequency side of the infrared peak is due to the limitation assumed for pointing and tracking accuracy of the space vehicle transmitting aperture.

3.2.2 Ground-based Receiver, Clear Weather

Similar maxima are indicated for a ground-based receiver under clear weather conditions. The performance at microwave frequencies, however, is shown as increasing out to about 10 GHz where atmospheric distortions tend to limit aperture size. This assumes the possibility of a single-element receiving antenna having an effective

diameter of 100 meters with a fabrication tolerance of about $\sigma/D = 2 \times 10^{-5}$ or, alternatively, the employment of an equivalent antenna array. Note that where information rates of the order of 10^6 bits per second are considered, an array large enough to provide this effective aperture at low elevation angles will require compensation for variation not only of the carrier phase but also of the signal delay across the aperture.

The performance peak in the infrared is lowered and shifted slightly to longer wavelengths by atmospheric distortion effects, while transmission in the submillimeter and far infrared region is effectively cancelled by atmospheric absorption. Smaller subsidiary peaks appear, however, at the 30 and 90 GHz "windows" in the millimeter regions.

For comparison the expected performance of two noncoherent optical systems is included. Because the particular detection mechanisms for these systems vary, it is not convenient to show a general functional dependence of performance on frequency.

3.2.3 Ground-based Receiver, Poor Weather

When the restrictions imposed by clouds and rain are considered the familiar single performance maximum appears in the microwave region at about 3 GHz. Undue significance should not be given to the exact position of the peak as shown, since it is determined more by the somewhat arbitrary choice of limits taken for transmitter and receiver antenna dimensions than by the more fundamental restrictions set by antenna noise temperature and atmospheric transmission.

3.3 SYSTEM CONFIGURATIONS

The curves of Figure 3-4 provide an indication of the relative performance to be expected of deep-space communication links operating in the various regions of the frequency spectrum. It is evident that a major hindrance to the improvement promised at higher frequencies is presented by the attenuating and distorting characteristics of the atmosphere.

Space missions which are likely to require the wide signal bandwidths considered in this study are those involving real-time transmission of data and hence an essentially continuous communications capability. (Such a requirement can be expected for a flyby mission, a landing mission, or a mission requiring voice communication.) At microwave frequencies continuous coverage can be achieved under all weather conditions, and a direct spacecraft-to-earth link is appropriate. However, at optical frequencies clouds may cut off a ground site completely. Thus, if an effectively continuous communication link is to be established between a spacecraft and earth within an acceptable atmospheric loss, alternative transmission configurations must be considered for wavelengths shorter than microwave. There seems little reason for serious consideration of a millimeter-wave system except possibly for a space vehicle to satellite link if a breakthrough in receiver technology should occur. For a ground receiver, performance at 30 GHz comparable to that at 3 GHz would require an equivalent antenna area with means for the correlation of atmospheric phase distortions over the aperture, or an order of magnitude increase in other system parameters. At 94 GHz ground receiver performance is down by another two orders of magnitude.

Potential microwave and optical system configurations are discussed briefly below with appropriate specific system parameters given in Section 3.4.

3.3.1 Direct Microwave Link

In the vicinity of 3 GHz atmospheric losses and restrictions on potential performance are small. At the same time much larger apertures can be reasonably contemplated for ground-based than for satellite-borne receivers. Thus a direct spacecraft-to-earth link is the obvious choice for a microwave system. Suitable components for such a system have been under progressive development for a considerable period of time, and only modest advancement of the state of the art should be required to permit information rates of the order of 10^6 bits per second. Note that for the extrapolated antenna gains

considered, a transmitted power of 10 watts from the space vehicle would be sufficient to meet this performance. Indeed the penalty paid for a microwave system is in the dimensions of the transmitting and receiving antennas required to achieve the desired performance at moderate transmitted-power levels.

3.3.2 Direct Optical Link

To provide an essentially continuous direct optical link between a spacecraft and the earth, multiple ground stations would evidently be necessary (i.e., several times the number needed to provide the requisite angular coverage). The number of additional stations required will depend on the statistical weather conditions at the specific sites available. The availability of a sufficient number of good sites could in principle result in an overall cost saving due to the smaller installation required for an optical receiver.

Coherent 10-Micron System. For coherent reception the CO_2 laser wavelength at 10.6 microns lies very near the performance peak in the infrared and with its high efficiency is the obvious choice for a coherent optical system. However, it is evident that multi-element apertures will be required to provide the levels of performance to which this study is directed. At 10.6 microns the product (S/N) (B) for each element is more than sufficient to permit phase-locking of the local oscillator to follow the low-frequency signal-phase distortions due to the atmosphere. The elementary signals can then be correlated at the heterodyne difference frequency to provide a useful signal-to-noise ratio at the information bandwidth.

A total effective aperture diameter of about 2 meters, comprising approximately 33 elements would be required to equal the performance of the 3 GHz microwave system considered (100 meter receiving antenna diameter). The number of elements could be reduced somewhat by the employment of larger but less efficient elemental apertures. The implementation of such a system would not be trivial, particularly when a goal of 10^8 bits per second for an optical system is sought, and substantial increases in other parameter values must be considered for high data-rate links.

Noncoherent 10-Micron System. Noncoherent detection at 10.6 microns is limited by thermal noise at the detector output, the sensitivity of the detector in this mode being less by about three orders of magnitude than in the coherent mode. (Because the limiting noise is independent of the signal, the actual degradation factor depends on the square root of the signal-to-noise ratio, which is taken here to be 10.) Inasmuch as atmospheric phase correlation lengths do not restrict the elemental aperture area, a single large "photon bucket" may be used. However, practical limitations on fabrication tolerances and detector dimensions will set an upper limit on noncoherent aperture diameters.

If a σ/D ratio comparable with that postulated for the 100-meter microwave antenna could be achieved for an optical reflector, diameters as large as 50 meters, sufficient to match the performance of the microwave link, might be feasible. In spite of the reduced complexity, the much greater aperture area required for noncoherent as opposed to coherent reception at 10.6 microns makes this system unattractive.

Noncoherent 0.5-Micron System. The communication link performance as indicated in Figure 3-4 discourages any present interest in coherent detection in the visible region of the spectrum for a ground-based system. However, the situation is somewhat improved for noncoherent detection. For a system operating at approximately 0.5 microns a photomultiplier detector can be used to provide essentially noise-free quantum detection. As compared with coherent detection at 10 microns, detection sensitivity will be down by a factor of two due to the higher quantum noise limit for noncoherent detection, another factor of about 2-1/2 to account for the lower quantum efficiency of the photoemissive detector surface, and a further factor of 20 due to the increase in quantum noise level with frequency. Sky background illumination will cause an additional degradation depending on receiver field of view and aperture size. For an optimized system this might be as low as 3db. As compared with a 10.6 micron coherent receiver, then, the total loss in sensitivity indicated is less than that for a noncoherent receiver at 10 microns.

However, a system at 0.5 micron suffers from one major disadvantage: presently known lasers in this region of the spectrum have efficiencies of the order of 0.1 percent. Thus, barring discovery of a new laser material, overall system performance would be over an order of magnitude below the low level anticipated for a noncoherent carbon dioxide (10.6 micron) system. While direct solar-pumping of the laser might give some improvement in overall efficiency, the improvement would not be sufficient to overcome the basic deficiency of the laser.

3.3.3 Satellite Optical Link

Another method of taking advantage of the increased aperture gains at optical wavelengths, while avoiding blackout due to poor weather conditions, is the use of a satellite relay station. The spacecraft-to-satellite link would take advantage of the higher aperture gains available at optical wavelengths while a short relay microwave link to earth would obviate atmospheric losses. Except for a few limited geometries a single synchronous satellite could provide a continuous link with the spacecraft so that a single satellite receiver and ground station might suffice for the full system. The cost of a satellite receiver must therefore be weighed against the cost (including maintenance) of three or four large-aperture ground-based microwave receivers, or of a greater number of relatively small-aperture, ground-based optical receivers.

Because of the two orders of magnitude better efficiency of the CO_2 laser over any other candidate laser presently known, 10.6 microns is the only wavelength worth considering at this time for an optical, spacecraft-to-satellite link. As indicated in Figure 3-4, performance of a coherent system with one-meter transmitting and receiving apertures falls slightly below performance of the 3 GHz spacecraft-to-earth microwave link. However, if effective aperture diameters are increased slightly to 1.2 meters (still within the pointing error limitation assumed), comparable performance results. Other than by an increase

in transmitted power a further substantial increase in performance can only be achieved by an improvement in pointing and tracking accuracy, coupled with either larger aperture dimensions or a higher transmitting frequency. The potential increase in performance is inversely proportional to the square of the effective pointing error for both transmitter and receiver. Thus the development of accurate pointing and tracking techniques is critical to the development of an optical space communications system.

For comparable performance, the aperture diameter of a noncoherent receiver would be about 30 times that for the coherent receiver and therefore unsuitable for a satellite relay station.

3.4 COMPARISON OF CANDIDATE SYSTEMS

Based on performance goals in terms of information rate, of 10^6 bits per second for a microwave link and 10^8 bits per second for an optical link, several specific candidate systems are compared in Table 3-4. A signal-to-noise ratio of 10 is adequate to give an acceptably low error rate for a suitable choice of modulation code; thus the performance goals set correspond to signal-to-noise bandwidth products of 70 and 90 db, respectively, for the two regions of the spectrum.

Parameter values are chosen generally within the practical limitations previously described with system and spreading losses as before. The increase in performance over present and planned systems as required to meet the goals is distributed among three system parameters: the two aperture areas and the transmitted power, the equivalent noise performance having already been extrapolated close to the fundamental quantum limit or to a reasonable temperature limit. The specific apportionment of these improvement factors is based on estimates of both development and engineering effort and costs required for their achievement. While a program is in progress to treat this problem rigorously, it represents a major undertaking and pertinent results are not available at this time to verify or modify the estimates made here.

Two optical configurations are shown in the table: a direct communication link between spacecraft and earth and a link employing a satellite relay station. Parameters for both configurations are given. An eventual choice between the two or between microwave and optical links must depend on rigorous systems and cost analyses and will certainly be influenced by the results of current and future research programs.

3.4.1 Direct Microwave Link at 3GHz

A 70 db information-rate parameter, $R_o = (S/N)(B)$, is achieved for the direct microwave link operating at a nominal frequency of 3GHz by postulating improved performance over the present state of the art for each of the three parameters: receiving antenna, transmitting antenna, and spacecraft transmitter. This represents an overall performance improvement of 19 db over the communications link projected for the Voyager mission. The increased requirements are apportioned among the various components as follows:

- Ground based antenna gain +9 db
- Spacecraft antenna gain +7 db
- Spacecraft transmitter power +3 db

The proposed receiver antenna gain of 70 db requires an aperture corresponding to a circular antenna diameter of 135 meters (55 percent efficiency) or about 440 feet at 3GHz. (At 2.3 GHz the required diameter would increase to about 580 feet.) The difficulties of implementing such an antenna in the form of a single dish are considered prohibitory. As discussed in Section 4.3.5, therefore, a distributed array appears to be the most promising solution. If a 70-percent aperture efficiency is assumed, a rectangular array capable of effective operation at zenith angles up to 30 degrees latitude and 60 degrees longitude would have dimensions of about 400 x 700 feet at 3GHz (or about 500 x 900 feet at 2.3 GHz). These dimensions are within expected atmospheric correlation lengths, so that no measures are needed to

compensate for atmospheric distortions. Operation is limited to a maximum zenith angle of 60 degrees for the array to avoid excessive and inefficient aperture dimensions; however, this will increase the minimum number of receiver sites required for continuous coverage from three to four. Note that at the reduced zenith angle atmospheric background noise will be decreased so that an operating frequency closer to 5GHz may be favored for the microwave link.

A gain of 39 db for the spacecraft transmitting antenna corresponds to a diameter of about 4 meters. Since weight added to the spacecraft entails a more than proportionate increase in booster weight, a considerably lower performance increase would normally be asked of the spacecraft component than of the ground-based antenna. However techniques currently being developed for the deployment of antennas after the boost phase of flight permit the consideration of larger apertures at higher performance to weight ratios. Thus an improvement close to that for the ground antenna is postulated for the spacecraft antenna.

A nominal 3 db advance is allocated to transmitter power. An increase in transmitter power involves a similar increase in spacecraft power supply and heat-dissipation capabilities and has, therefore, a substantial effect on overall vehicle weight and booster requirements. In as much as this weight penalty is unlikely to be reduced by any major improvement in the efficiencies associated with these various processes, the allowance for transmitter power has been increased to just 100 watts.

3.4.2 Direct Optical Link at 10.6 Microns

An information-rate parameter of 90 db is assumed as the goal for an optical deep-space communication system in view of the greater potential performance promised and in justification of the development of optical communication techniques.

For coherent reception of a signal arriving at a zenith angle of 70 degrees, receiver aperture of only a half meter (50 percent aperture illumination efficiency) would be expected to suffer a 3 db loss in

aperture gain due to typical atmospheric distortions of the signal wavefront. Careful selection of the site and perhaps tower-mounting of the receiver should permit doubling of the aperture diameter for the same overall aperture efficiency, giving a 6 db enhancement in the signal. A further 3 db can be gained by increasing the aperture diameter an additional factor of two at the expense of a factor of two loss in efficiency. For greater diameters, overall aperture efficiency drops rapidly and there is little advantage to further increases. A single-element receiver aperture is therefore limited to about a 2-meter diameter corresponding to a nominal gain of 112 db but with a 6 db loss (included among the fixed losses in Table 2) due to atmospheric turbulence.

The gain of the transmitter aperture is limited by the accuracy with which the transmitter can be pointed at the earth station. For an effective pointing accuracy of five microradians or a minimum allowable beamwidth of ten microradians the gain is limited to 112 db corresponding again to a 2-meter aperture diameter (50 percent efficient).

For a 20-db improvement in performance over the microwave link it seems appropriate to ask for an increase in transmitted power of the order of 7 db, account being taken of the fact that an increase in spacecraft weight must be paid for by a disproportionate increase in booster thrust. This would give a transmitted power of 500 watts corresponding to a raw power requirement of about 3.5 kilowatts, which is not out of line with spacecraft capabilities estimated for 1980.

The overall performance, in terms of the information-rate parameter, for a single-element ground-based receiver is therefore 85 db. An additional 5 db is required to meet the 90 db goal. The deficit can be made up in principle by an improvement in spacecraft pointing accuracy from five to three microradians but at the expense of a corresponding increase in aperture diameter. Alternatively, a three-element receiver aperture could be used with individual aperture elements phase-locked to the incoming signal to compensate for the atmospheric phase distortion, and with signal correlation accomplished at the IF frequency. The latter case is shown in Table 3-4.

Configuration:		Direct Link				Satellite Relay	
Frequency/Wavelength:		3 GHz		10.6 micron			
Receiver							
Aperture diameter (rectangular array)	D_R	135 m (400 x 700 ft)		2 m		2 m	
Aperture efficiency (array)		0.5 (0.7)		0.5		0.5	
Aperture gain	G_R		70 db		112 db (117 db)*		112 db
Transmitter							
Aperture diameter	D_T	4 m		2 m		2 m	
Aperture efficiency		0.5		0.5		0.5	
Aperture gain	G_T		39 db (49 db)+		112 db		112 db
Power output	P_T	100 w (1000 w)+	20 dbw (30 dbw)+	500 w	27 dbw	500 w	27 dbw
Performance Limitations							
Noise spectral density**	$\frac{1}{kT_e}, \frac{\eta_q}{hf}$		214 dbw		191 dbw		191 dbw
Spreading loss (1 AU)	$\left(\frac{\lambda}{4\pi R}\right)^2$		-266 db		-345 db		-345 db
Transmission loss**	$1/L_f$		-0.2 db		-2.2 db		0 db
Fixed losses**	$1/L_o$		-7 db		-10 db		-7 db
System Performance	(S/N) (B)		70 db (90 db)+		85 db (90 db)*		90 db

*Three-element aperture
**Reciprocal
+Alternate configuration making use of extensible spacecraft antenna and spacecraft power comparable to optical case to achieve 10^8 bits-per-second data rate.

Table 3-4. Deep-space communication system parameters.

3.4.3 Satellite Relay Link at 10.6 Microns

For a satellite receiver, aperture diameter is limited as for the transmitter aperture by achievable pointing accuracies. For a receiver, however, the pointing problem reduces to a simple closed-loop angle-tracking problem internal to the receiver. Thus considerably smaller pointing errors can be anticipated for the receiver than for the transmitter where the detected error angle must be translated to a predicted pointing angle for the transmitter to overcome such problems as boresight misalignment and lead angle.

If an achievable rms error of 1.5 microradians is assumed for the receiver tracking angle, a gain of 122 db should be allowable for the receiver aperture. This corresponds, however to a diameter of about six meters and requires optical-quality fabrication tolerances of the order of $\sigma/D = 10^{-7}$. The fact that such an aperture exceeds the dimensions of the Mt. Palomar telescope should not be used to disconnect entirely the possibility since, for a monochromatic radiation, focusing techniques such as the use of Fresnel zone plates might be developed to permit light-weight, space-deployable apertures.

A more predictable solution to the problem, however, would be a receiver aperture of two meters, equal to the spacecraft transmitter aperture, and having a gain of 112 db.

A transmitted power of 500 watts, similar to that for the direct spacecraft-to-earth link is then required to meet the desired performance goal of 90 db.

The choice between a satellite-relay and a direct optical communication link reduces to an evaluation of the costs of a satellite receiver (one or at most two required) as opposed to multiple ground-based receivers in widely-dispersed locations.

3.5 REFERENCES

1. K. L. Brinkman and L. S. Stokes, Deep Space Communication: System Trade-Off Parameters, IEEE Winter Convention on Aerospace Electronic Systems, Los Angeles, February 2-4, 1966.

2. John Ruze, Physical Limitations on Antennas, TR No. 248, Research Laboratory of Electronics, MIT, ASTIA/AD 62351, October 1952.

3. See Reference 45, Section 2.6.

4. P. D. Potter, W. D. Merrick, and A. C. Ludwig, Large Antenna Apertures and Arrays for Deep Space Communications, TR No. 32-848, Jet Propulsion Laboratory (NASA), CIT, November 1, 1965.

5. See Reference 56, Section 2.6.

6. Research in Space Science, Special Report No. 176, Smithsonian Institution Astrophysics Observatory, 17 May 1965.

4.0 RADIO FREQUENCY TECHNOLOGY

The present state of RF technology is well advanced. Relatively efficient transmitting power sources and radiation detectors exist throughout the microwave and millimeter wave regions although little consideration has been given to the submillimeter region so far. (The severe atmospheric attenuation in the submillimeter range restricts useful terrestrial applications, and no advantage appears to be offered over operation in the far IR for space applications.) While small improvements in transmitter and receiver efficiencies and perhaps an order of magnitude increase in available spacecraft supply power can be expected over the next decade, a substantial increase in the channel capacity of deep space communication links must come from higher gain antennas.

The state of the art and the practical limitations of transmitting power sources, detectors, and front-end components are discussed first. This is followed by a discussion of antennas including the various types, the performance limitations that apply, and appropriate candidates for deep space communications systems.

4.1 RADIO FREQUENCY SOURCES[1]

4.1.1 Microwave Sources

A number of tube types are available in the microwave region of the spectrum (1-30 GHz); these are summarized in Table 4-1. The types of most interest are:

- The klystron, a well-developed and reliable tube
- The traveling wave tube (TWT), an inherently high average power device
- Crossed-field devices, high in efficiency and light in weight, which include the conventional magnetron, the amplitron, and the many linear beam-type magnetron amplifiers

The significant characteristics of these three types are summarized in Tables 4-2, -3, and -4, respectively.

			Interaction Circuit					
			Backward Wave		Forward Wave		Standing Wave, cavity	
			Oscillator	Amplifier	Oscillator	Amplifier	Oscillator	Amplifier
Electric-Magnetic Fields	Crossed Field (M-type)	Injected Beam	M-BWO M-Carcinotron	M-BWA Bitermitron	TPOM	CFA TPOM Bimatron		
		Continuous Cathode	Stabilotron	Amplitron CFA	VTM	FWA-CFA Dematron	Magnetron	Circlotron
	Linear (Single or Parallel) Field (O-type)	Injected	O-BWO O-Carcinotron	O-BWO		TWT TPO	Klystron Reflex Klystron Monofier Monotron	Klystron

Table 4-1. Type classification of microwave tubes - oscillators and amplifiers.

ADVANTAGES

1. Single envelope is practical for gains up to 30 db, depending on output power levels.
2. Much development experience exists both at high power (megawatts) and lower drive levels (tens of kws).
3. Bandwidths to 10 percent have been achieved in high-power units; 5 percent or less is more realistic at low power.
4. Klystrons can be made to operate with dc beam supplies by employing switching or modulating anodes.
5. Klystrons still offer higher power than any other tube type.
6. Focusing is normally performed by solenoids.
7. More design and development experience exists on this type of beam device than on any other.
8. Higher perveances have been achieved than in TWTs. This permits lower beam voltages for equivalent output power.
9. System reliability is enhanced in cases where 30 db gain is sufficient, since a single amplifier stage is adequate.
10. Cooling techniques are known and optional; air and liquid are common.
11. Since the klystron is a unidirectional device, operation into a mismatched load is possible without isolation, depending on system limits of power and phase shift.
12. More recently developed electrostatically focused klystrons offer reduced size and weight.
13. Shorter electrical length per unit gain than the TWT, thus suffering less voltage-phase sensitivity.

DISADVANTAGES

1. Longer electrical length per db of gain as compared to cross-field devices.
2. Gains are about half those obtained with TWTs.
3. Klystrons require a filamentary cathode and gun for operation, thus more electrode voltages are needed.
4. Higher beam voltages required for a given output power than in crossed-field devices.
5. Bandwidth limitations are severe at low power (tens kws); 2 to 5 percent is typical.
6. Good voltage regulation may be required for acceptable phase stability.
7. Efficiencies of 25 to 35 percent are typical.
8. Structure offers moderate filtering; may require separate high power filters at an additional cost to the system.
9. Reliability not reasoned to be as good as cold cathode devices operated without gun.

Table 4-2. Klystron characteristics.

ADVANTAGES	DISADVANTAGES
1. Single envelope is practical for gains to about 60 db.	1. Long electrical length per db of gain in comparison to other tubes.
2. Much development experience exists both at high power levels (megawatts) and lower drive levels.	2. Phase sensitivity of RF output is very dependent on beam voltage, because the device operates on the principle of synchronism between electron velocity and RF velocity on the slow-wave structure.
3. Bandwidths beyond 10 percent are common in today's TWTs.	3. TWTs require a filamentary cathode and gun for operation – more electrode voltages are thus needed.
4. Can operate with dc beam supplies by employing switching or modulating anodes. Grid control is possible at power levels below 10 kw.	4. Low perveance of TWTs requires higher beam voltages for a given power output compared to other devices. High perveances involve hollow beams, whose increased current density may cause focusing difficulties and cathode loading problems.
5. Long length, small-diameter form factor is suitable for phased arrays limited to tube diameters of less than $\lambda/2$.	5. High gains usually require solenoid focusing (or a recently proposed magnetic matrix focusing technique). Solenoids for high power tubes are large, require substantial power, and create packaging and distribution problems.
6. Several types of slow-wave structures can be cascaded in a single envelope to optimize the RF coupling design.	6. Acceptable phase stability calls for good beam voltage regulation.
7. Focusing can be accomplished by electromagnets, permanent magnets, or electrostatically, depending on system requirements.	7. Efficiency is $\approx$ 25 percent without depressed collectors. Depressed collectors raise efficiency by $\approx$ 10 percent.
8. Reliability is enhanced by single-stage operation.	8. Electrical structure offers very little filtering. Separate high-power filters may be needed.
9. Depressed-collector techniques ease regulation requirements on high-current beam supply.	
10. Cooling techniques are known and operational; air and liquid are common.	
11. Severed or attenuated slow-wave structures enable operation without circulators into a mismatched load, depending on the limits of phase tolerance.	
12. No doubt exists about development and mass production of TWTs.	

Table 4-3. Traveling wave tube characteristics.

4.1.2 Millimeter Sources

Most of the successful attempts to develop sources in the millimeter wave region (30-300 Ghz) have involved extensions of microwave techniques. Since the slow-wave circuit must have dimensions comparable with the wavelength, however, the problems of electron control and thermal dissipation become formidable. These difficulties are overcome more successfully in linear beam, or O-type distributed interaction devices.

ADVANTAGES	DISADVANTAGES
1. Gains to 45 db possible with injected beam variety; 10 percent bandwidth typical.	1. Limited gain of 10 to 15 db for distributed emission and types.
2. Good phase stability; fairly independent of anode voltages, since velocity synchronism is not an operating requisite.	2. Has very limited dynamic range of output power for a particular anode voltage and varying input power.
3. Distributed emission type has shortest electrical length per unit gain of all tubes.	3. Crossed-field tubes are bidirectional. Any reflections into the output will return directly to the driver or input circuit. A circulator or isolator is thus necessary.
4. Can operate with a cold cathode; electron gun or filaments are avoided.	4. Injected beam types require very good regulation of sole voltage for good phase stability. Typical values are 20 degrees for a 1 percent change in sole voltage.
5. Reasoned to have long life compared to filamentary devices.	5. In distributed emission types it is difficult to initiate emission.
6. Small in size and relatively light in weight.	6. Perveance of injected beam types is comparable to that of TWTs, thus requiring higher beam voltages for a given output power.
7. Distributed-emission type provides the lowest anode voltage for a given power output. This simplifies the power distribution system.	
8. Air or liquid cooling is acceptable.	
9. Highest microwave tube efficiencies: Injected beam – 35 to 50 percent Distributed emission – 45 to 65 percent	
10. Operated at saturation to give a constant output independent of input power variation; makes a good limiter.	
11. Can be used in a duplexed system with receiver and duplexer on the low input side.	
12. Structure is ideally suited for mass production under tight mechanical tolerance control.	

Table 4-4. Crossed-field tube characteristics.

Available power levels for millimeter-wave tubes have increased by three orders of magnitude since 1960, and efficiencies have been increased to be competitive with microwave sources. Power levels of the prominent tube lines supplied by various manufacturers are shown in Figure 4-1 for output ratings in excess of one watt.

4.1.3 Submillimeter Sources

Conventional RF techniques become impractical in the design of devices for operation at frequencies above 300 Ghz. At present the most effective means of producing power at these frequencies involves harmonic generation from millimeter sources. The problem then becomes the development of a nonlinear device to accept large input powers. Notable success has been achieved with high-pressure gas

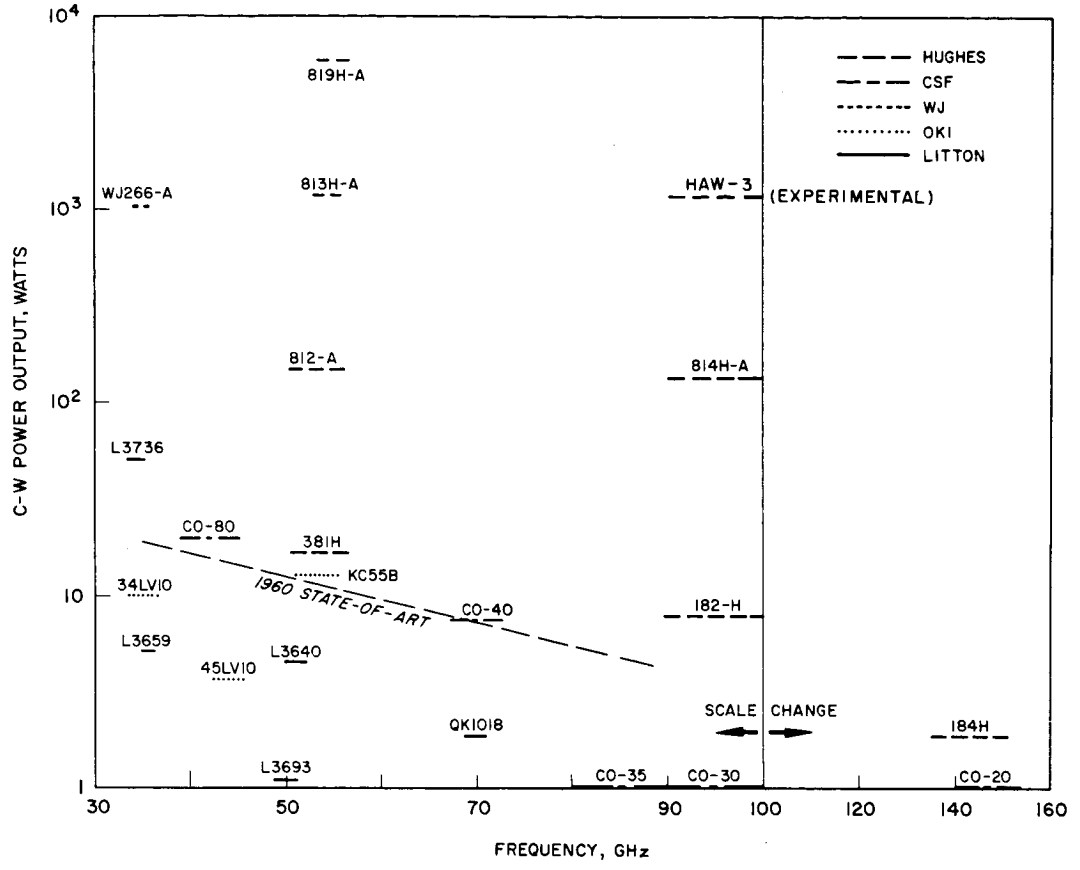


Figure 4-1. Power characteristics of available high-power CW sources.

discharge plasmas in this application, and there appears to be considerable promise for further development. Maser (or SMASER, the first two letters standing for "submillimeter") action has also been demonstrated[2] in this region for laboratory devices having very low output powers.

4.1.4 Solid-State Sources

Considerable attention is currently being given to the use of solid-state transit-time devices as oscillators and amplifiers in the microwave region. There are two types of transit-time devices: those utilizing the Gunn effect[3] and those based on avalanche breakdown.

At present CW output power levels are in the 1-100 milliwatt range with efficiencies of only a few percent. However, the development of techniques for correlation of the coherent output of multiple oscillators could provide the basis for a distributed, light-weight transmitting source.

4.1.5 Burden Considerations

Since device dimensions generally scale with wavelength, there is a corresponding reduction in allowable fabrication tolerances for efficient operation. The required manufacturing tolerances for traveling-wave tubes are shown, for example, in Figure 4-2 as a function of frequency. Thus manufacturing capabilities and associated costs set a limit to the upper frequency range of mechanically structured devices.

Power dissipation in the device imposes another restriction on performance. The rate of dissipation of RF-generated heat in the power source is proportional to the area available for heat conduction. Hence the power obtainable from a single device (for a given efficiency) will vary, in general, as the inverse square of the frequency as shown in Figure 4-3 for a TWT. This does not of itself represent a burden on the system inasmuch as multiple units may be used as the transmitting source. The weight per unit does not drop as fast in practice, however, as the maximum output power, and as a result there is a decrease in power to weight ratio (P/W) with frequency. This trend is shown in Figure 4-4 based on state-of-the-art data.

Also plotted in Figure 4-4 is the function η^2/λ^2 and the product:

$$M_{SB} = (\eta^2/\lambda^2)\,(P/W)$$

where η is the efficiency of the device and λ its operating wavelength. M_{SB} represents a figure of merit for a spaceborne transmitter, neglecting subsequent transmission and reception considerations. The efficiency, η, appears as the square in this figure of merit since it

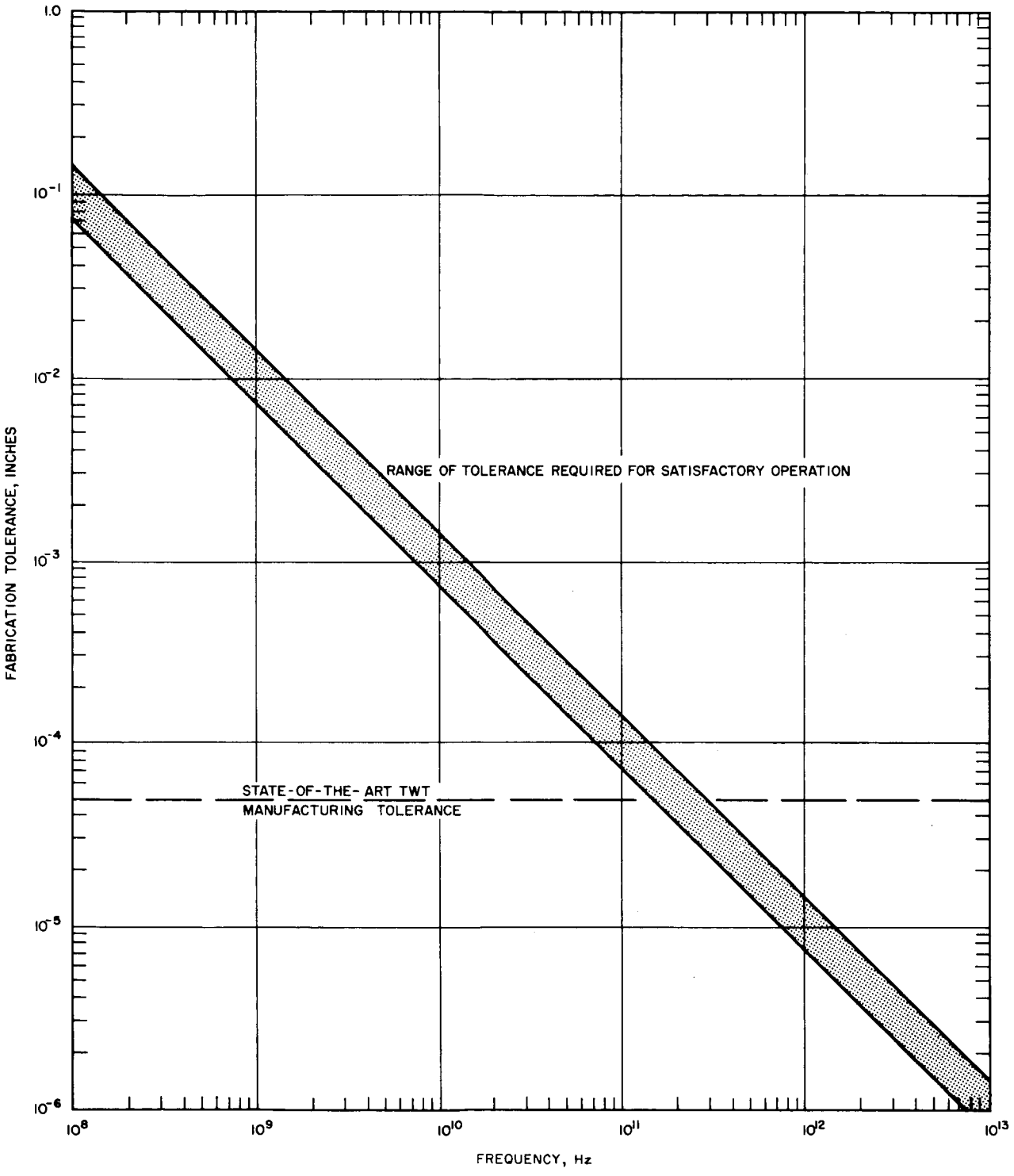


Figure 4-2. Required manufacturing tolerances for TWTs as a function of frequency.

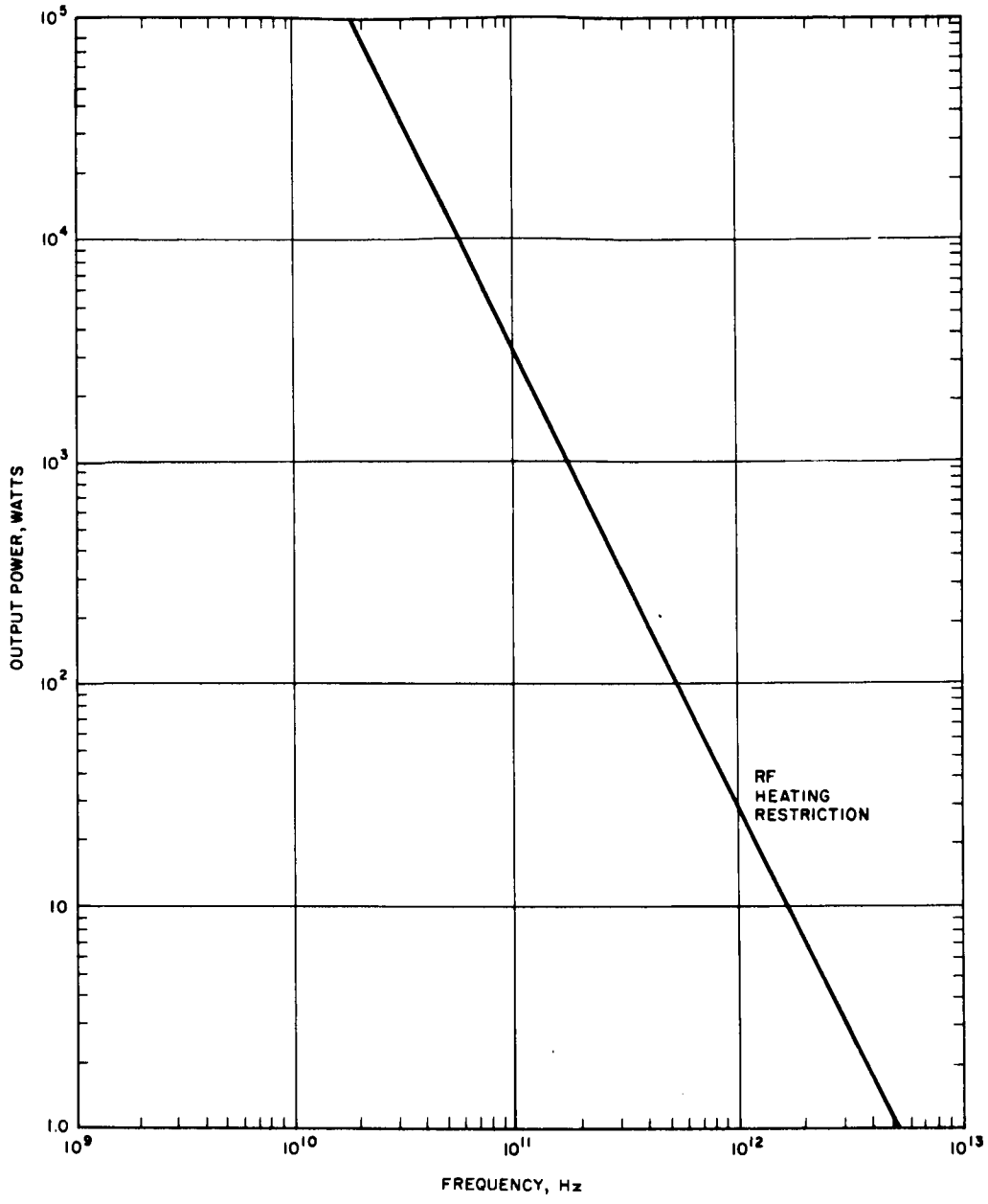


Figure 4-3. Power limitations for a single TWT.

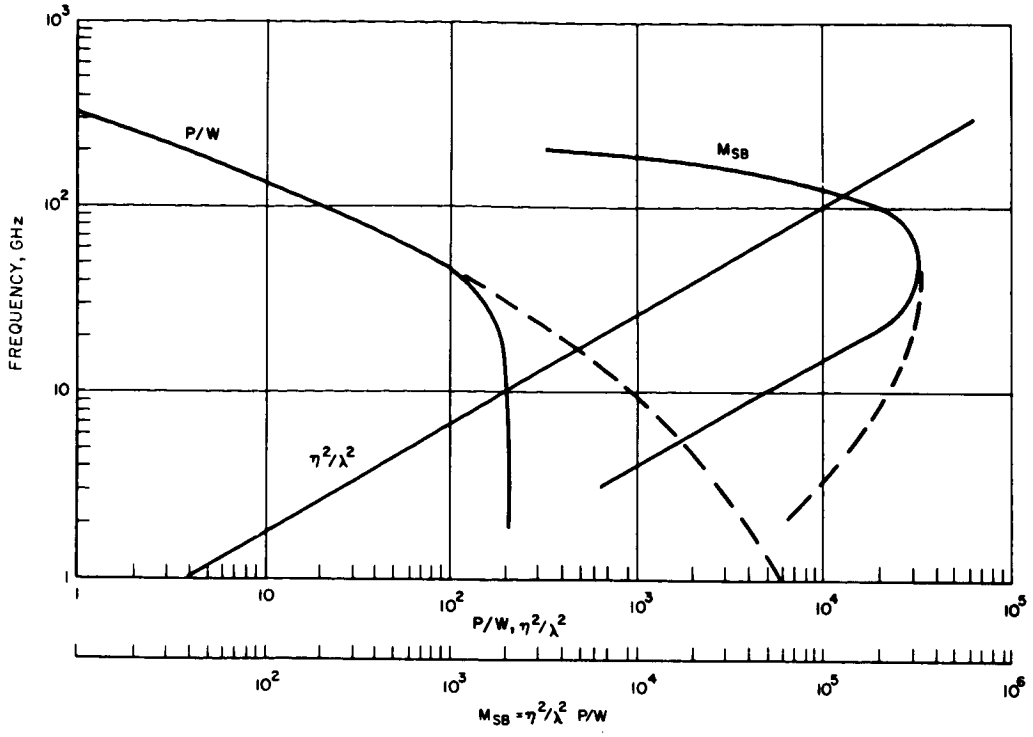


Figure 4-4. Power to weight ratio and figure of merit for spaceborne CW transmitter.

affects the spacecraft requirements for both supply power and heat dissipation. The λ^2 factor represents the improvement in antenna gain for a fixed allowable antenna size.

4.2 RADIO FREQUENCY DETECTORS[1]

Detector or receiver sensitivity is conventionally characterized by the effective noise temperature, which is a measure of the receiver noise referred back to its input terminals, or by the noise figure, defined as

$$NF = \frac{S_{in}/N_{in}}{S_{out}/N_{out}}$$

The noise temperature, T_e, is related to the noise figure by the expression

$$T_e = (NF-1)\, T_o$$

where T_o is a standard temperature taken to be 290° K.

4. 2. 1 Micro/Millimeter-Wave Detectors

The more sensitive detection devices in this frequency range are the conventional crystal mixers, parametric amplifiers, and masers. Harmonic mixers are often employed at the high end of the range.

Crystal mixers are normally characterized by their conversion loss. The noise figure of the mixer is then

$$NF = L_C (NF_{IF} + N_R - 1)$$

where L_C is the conversion loss, NF_{IF} is the noise figure of the IF amplifier, and N_R is the crystal noise ratio. The conversion loss is defined as the ratio of RF input power to the measured IF output power at the mixer. The crystal noise ratio, N_R, is the ratio of noise power developed by the crystal to the thermal or Johnson noise of an equivalent resistance at 290° K and is typically about 2 in a well-designed system. Estimates of best available noise performance from crystal mixers are summarized in Table 4-5.

Parametric amplifiers are generally available in the microwave range, but in the millimeter region availability is restricted to experimental models. Diode cutoff frequencies, associated with spreading resistances and junction capacitances, prevent the practical extension of operation beyond 100 GHz except by direct insertion of the diode into a cavity.

Masers are available in the microwave range. As is the case for parametric amplifiers, operation at millimeter wavelengths has been restricted to experimental models. Representative values of the best noise performance at several frequencies are given in Table 4-7.

Hot carrier detectors are based on the application of microwave power to a noninjecting point contact. As the majority carriers are excited, a temperature gradient between the point contact and broad contact is established, and a unidirectional thermoelectric voltage is generated with frequency following as high as 100 GHz. Work[4] on these detectors is in the experimental stage.

Frequency	Conversion Loss*	Noise Figure	Noise Temperature
10 GHz		6 db	870°K
35 GHz	5.5 db	10.3 db	2,800°K
60 GHz	6.5 db	11.3 db	3,600°K
94 GHz	8.2 db	12 db	4,300°K
140 GHz	9 db	13.8 db	6,700°K
200 GHz	19-20 db	25 db	91,000°K
300 GHz		35 db	910,000°K

*Where only conversion losses are quoted, IF noise figure and crystal noise ratio are taken to be 2.

Table 4-5. Noise performance estimates for crystal mixers.

Frequency	Noise Figure
1 GHz	0.8 db
3 GHz	1.3 db
9 GHz	2 db
14 GHz	3.5 db
94 GHz (estimated)	10 db

Table 4-6. State-of-the-art performance of parametric amplifiers.

Frequency	Noise Temperature
2 GHz	10-15°K
8 GHz	20-25°K
35 GHz	130°K
81.3 GHz	300°K
94 GHz	200°K

Table 4-7. State-of-the-art performance of masers.

4.2.2 Submillimeter-Wave Detectors

The mechanisms for detection of submillimeter radiation are based on either a thermal effect or a photoelectric effect. Since thermal response times are generally long (in the millisecond range), applications to high-data-rate communication systems are limited to detectors utilizing photoelectric mechanisms. In general, these devices must be cooled to reduce thermal lattice vibrations so that only electrons absorbing the low electron energy (1.2×10^{-3} ev for $\lambda = 1$ mm) are excited to the conduction band. Detection has been demonstrated using several semiconductor materials, and a superconductive detector based on electron tunneling has been proposed.[5] In addition, detection of submillimeter and millimeter waves by down-conversion to the microwave region has recently been proposed and analyzed.[6] The scheme utilizes an appropriate quantum energy-level system of at least three levels and is similar in operation to a three-level maser where the roles of the pump and signal are reversed. At present, however, definitive data are not generally available on detectors in the submillimeter region.

4.3 LOW NOISE PREAMPLIFIERS

The use of multiple-element receiving antenna arrays presents a special problem. Because of the losses of beam forming networks, phase shifters, and other phased array components, it is necessary to include some form of low noise amplification before the high loss component is encountered. The level at which the preamplification takes place and, thus, the number and the gain of preamplifiers depend on the quality of the other phased-array components. Most large phased array systems now in use or under study utilize one preamplifier per group of radiating elements or even one per radiating element. For the lowest temperature possible, an amplifier per element should be used. Considerations of cost, size, and system noise temperature requirements and the state of development of the other components will determine the optimum ratio of the number of preamplifiers to radiating elements. In any case, a large number of preamps is likely to be required.

At the present time there are five low-noise microwave amplifiers which may be considered for use in a large phased array deep space communication system. These are:

- Transistor Amplifier
- Tunnel Diode Amplifier (TDA)
- Traveling Wave Tube (TWT)
- Parametric Amplifier
- Maser

Brief discussions of the characteristics of these amplifiers are given in the following paragraphs:

4.3.1 Transistor Amplifiers

Microwave transistor amplifiers are relatively new devices which have promise of moderately low noise figures. At the present time noise temperatures of 200 to 625°K can be obtained at frequencies up to about 1 GHz with approximately 20 db gain. It is estimated that in ten years 120° to 170°K noise temperatures are likely at 2 GHz and feasible up to 15 GHz. Transistor amplifiers appear to have their most useful application as a second stage amplifier following an ultra low noise amplifier.

4.3.2 Tunnel Diode Amplifiers

The tunnel diode amplifier (TDA) is the simplest solid-state microwave amplifier and has moderate gain and noise characteristics. Noise temperatures range from about 360°K at 1 GHz to 520°K at 10 GHz using gallium antimonide diodes. Germanium diodes have about 1 db higher noise figures but are available for operation up to about 20 GHz. Single stage amplifiers normally provide about 17 db gain. However, stable gains as high as 30 db can be obtained by careful attention to temperature control and power supply stability. Bandwidths are more than adequate for space communication systems. Noise figures of room temperature TDA's are not expected to improve significantly. Such TDA's will, therefore, be of use mainly as second stage devices following ultra-low noise amplifiers.

The noise generated by a tunnel diode amplifier is largely caused by shot noise from the diode bias current. The magnitude of this noise contribution is determined by the characteristics of the material used and can be reduced by using low energy gap materials such as indium antimonide. Such materials must be operated at cryogenic temperatures. However, assuming that a suitable material can be found, a cryogenic TDA with a noise temperature of 20°K at 1 to 2 GHz would have a strong advantage over cooled paramps and masers in that no pump power would be required.

4.3.3 Traveling Wave Tubes

These devices offer high gain and moderately low noise figures. Noise temperatures in the 360 to 440°K range are presently obtainable over narrow bandwidths up to 2 GHz. It is unlikely that noise temperatures below about 225°K will be consistently obtained in the next decade. The two major noise sources in a TWT are beam shot noise and thermal noise from the attenuator. Present low noise TWT designs require complex anode structures to achieve space charge smoothing for shot noise reduction. Since no significant advances in space charge smoothing have been made since the late 1950's, it is not anticipated that major progress will be made within the next decade. Some improvement in noise temperatures can be expected by cooling the attenuator, but it does not appear that TWT's will be competitive with cooled paramps.

4.3.4 Parametric Amplifiers

Parametric amplifiers have demonstrated room temperature noise performance superior to that of transistor and tunnel diode amplifiers and cryogenic noise performance approaching that of the maser. In recent years the uncooled parametric amplifier has achieved a level of reliability that has permitted applications on a broad basis and in large numbers. Noise temperatures range from about 60°K at 1 GHz to 250°K at 10 GHz for well-designed narrow band amplifiers.[7] When cooled to 20°K, noise temperatures between 14°K at 1 GHz and 30°K at 10 GHz are possible with careful design using

presently available components. Narrow band gains as high as 30 db are possible using extremely stable temperature and pump power control. Major advances for the cooled paramp are likely to be in cost reduction and reliability particularly in the associated cryogenic equipment.

4.3.5 Masers

Masers find applications in special areas where the ultimate in low noise performance is either dictated by technical requirements or provides the most economical solution to the problem. The noise temperature of the maser itself is approximately that of its physical temperature, approximately 5°K. To this must be added the noise contribution of the section of input transmission line over which the temperature transition to room temperature is made. For frequencies in the range 1 to 20 GHz this contribution can be held to 5 to 10°K giving an overall maser noise temperature of 10 to 15°K. Gains of better than 30 db with bandwidth of 1 to 2 MHz or more are readily obtainable with a cavity maser. The major disadvantage of masers is that they must operate at a temperature of a few degrees Kelvin in order to provide sufficient gain. The complexity and cost of a cryogenic system increases rapidly as the temperature approaches 0°K.

Future improvements in the maser are also likely to be in the area of reliability and cost reduction.

4.3.6 Summary

A summary of the important properties of low noise amplifiers is given in Table 4-8. Of the devices surveyed, only the maser and helium cooled paramp can presently meet the low-noise temperature requirements for the down-link receiver in a deep-space communication system, and a system based on reasonable extensions of today's state of the art would require one of these two devices. The maser requires a physical temperature of 4.2°K or less while the paramp can provide adequate noise performance at physical temperatures up to 20°K. The cryogenic system for the paramp would be considerably less complex and less expensive. Barring the discovery of new types of ultra low

Amplifier Type	Gain Per Stage, db	Coolant Temperature °K	Relative Cost (1975-1980)	Relative Reliability (1975-1980)	Present Noise Temperatures (Amplifier Only), °K			Estimated Best Noise Temperature (1975-1980) (Amplifier Only), °K		
					Frequency, GHz			Frequency, GHz		
					1	2	10	1	2	10
Traveling Wave Tube	35	290	Medium to High	Medium	380°	400°	1340°	225°	250°	350°
Tunnel Diode Amplifier	20	290	Low	High	360°	380°	525°	310°	330°	480°
	20	20	Medium to Low	Medium	225°*	225°*	---	20†	25†	35°†
Transistor Amplifier	20	290	Low	High	200°	625°	---	75°	125°	170°
Parametric Amplifier	20	290	Medium	High	60°	80°	250°	30°	40°	130°
	20	20	High	Medium	18°	20°	50°	15°	15°	20°
Maser	30	4.3	Very High	Low	10-15°	10-15°	10-15°	12°	12°	12°

*Estimate based on use of indium antimonide tunnel diode at 77°K
†Feasibility not established

Table 4-8. Low noise preamplifier characteristics.

noise amplifiers, the helium cooled parametric amplifier operating at 20°K presently appears to provide the most economical solution. Transistor and tunnel diode amplifiers could be used as low noise second stage devices.

For the implementation of large receiving-antenna arrays, then, research should be directed toward cost reduction and mass production techniques on cooled parametric amplifiers. At the present time the future of the cooled tunnel diode amplifier is not certain. If a suitable material can be found, it is, at least in principle, possible to construct a tunnel diode amplifier operating at cryogenic temperatures with noise performance approaching that of the cooled paramp. If such a material can be found, the desirable features of the cooled TDA including basic simplicity, no RF pump power, and moderate cryogenic requirements will make it an attractive device for space communication systems.

4.4 ACTIVE RF COMPONENTS

The potential use of phased arrays for spacecraft transmitting antennas requires a means of adjusting the phase at each element. Conventional RF phase-shifters which are discussed under Antennas (Section 4.5) have certain disadvantages: loss, weight, size, and power consumption are some of the problems with these devices. Loss is doubly undesirable because of the RF heating and because of the cost in terms of weight, power, and reliability required for generation of the RF power.

As an alternative to RF phase shifters, the phase shift can be accomplished at a lower frequency and lower power level; the proper phase is then injected into a mixer for conversion to the desired RF frequency, and the RF energy is delivered, either directly or through an RF amplifier, to the radiating elements.

4.4.1 RF Amplifiers and High-Level Mixers

There are several RF amplifiers and high-level mixers which can be considered. The solid-state transistor amplifier and the traveling-wave tube (TWT) constitute two types. The TWT improves in efficiency

as the RF power rating is increased; the filament power remains rather constant irrespective of the RF power rating and thus accounts for a large part of the loss of efficiency at low levels. TWT's up to 30 percent efficiency have been space qualified with operating lives over 10,000 hours. A gradual, though limited increase in efficiency is anticipated with, perhaps, some decrease in weight as the technology improves. TWT's can be used in three capacities: (1) as the final RF amplifier in the transmitter, (2) as the amplifier in every radiating module (for every module one TWT), and (3) as a pump for high-level mixers.

The solid-state transistor RF amplifiers are in a state of flux. Below 2 GHz, 5 watts (at 600 MHz) are being obtained. At 2 GHz, the projection is for several watts with a 30-percent efficiency and a gain of about 30 db, if the expectations are realized at this frequency.[8,9] As technology improves, this frequency will be increased to perhaps 4 GHz with similar power. However, the MERA* engineers are pessimistic about the higher frequencies, say 7 or 8 GHz. One alternative would be the utilization of a frequency multiplier which is driven by a 2-GHz amplifier. The efficiency is still satisfactory, but phase as well as frequency is multiplied.

The other approach is to use high-level mixers (also called resistive or varactor up-converters). These devices are perhaps interim in the state-of-the-art between solid-state RF amplifiers and TWT amplifiers for modular units. Such a high-level mixer is similar to the low-level mixer except for the direction of power flow for some of the signal frequencies. It has three signal frequencies: RF input, RF output, and IF input. Other signal components are rejected by proper design. Either resistive or varactor mixing may be employed. The former is limited to low levels on the order of 10 mw. Varactors are capable of 100 mw output per pair.

The efficiencies of these devices are good but they suffer overall system loss since an RF pump is required. The RF pump power must

*Molecular Electronics for Radar Applications

be generated elsewhere, presently by TWT's and in the future perhaps by solid-state oscillators. The overall efficiency of varactor mixer and pump is perhaps 10 to 15 percent at best. Thus, the most promising devices in the long run seem to be the solid-state RF amplifiers for distributed amplification. Table 4-9 summarizes the characteristics of various RF amplifiers and high-level mixers as available today.

It should be pointed out that phase and gain tracking must be maintained between all amplifier modules. The most critical type of amplifier in this regard is the TWT since it has many electrical degrees between input and output. However, excellent gain and phase tracking have been achieved from tube to tube for tubes specifically designed for phase and gain tracking. From 8 to 12 GHz, deviation from true time delays is 15° and 1 db rms.

4.4.2 Use of Multiplier Chains for Power Output Stages

One of the problem areas in self-steering arrays at higher frequencies is the non-availability of efficient, medium-power, RF amplifiers in the 100 mw range for use as final stage power amplifiers. Transistor power amplifiers are presently suitable only up to about 2 GHz. For frequencies between 2 and 10 GHz the most practical RF power source presently available seems to be the varactor multiplier. Traveling-wave tubes become competitive in this frequency range only when amplifier output powers of one or more watts are desired.

One class of self-phasing system uses frequency multipliers in a novel way. In place of final RF power amplifiers, this kind of array utilizes harmonic multipliers preceded by transistor power amplifiers at lower frequencies. In many versions of systems based on this basic harmonic multiplier approach, the incoming pilot and/or information signals are divided in both frequency and phase so that after the signals are multiplied back to the higher frequency for transmit, the same bandwidth is retained as was present before they were divided down. It should be mentioned that this scheme is generally restricted to FM

Type	Frequency Range (GHz)	Power Output	Power Input	Efficiency	Weight	Comments
Traveling-wave Tube	2-10 at a selected frequency	200 mw 1 watt	gain 40 db	8% 10-15%	12 ounces 14 ounces	Efficiency limited by heater power; use as module amplifier.
Traveling-wave Tube	4-8 at a selected frequency	10 watts	gain 30 db	30%	40 ounces 2-1/2 pounds	HAC Syncom tubes-space qualified.
Traveling-wave Tube	2-10 at a selected frequency	35 watts	gain 40 db	28%	42 ounces 2-1/2 pounds	Example chosen Watkins-Johnson tube WJ-231-6 at 7 GHz-space qualified.
High-Level Mixer-Resistive	2-10	10 mw maximum (difficult to achieve)	100 mw RF 400 mw IF	10 db minimum conversion loss RF to RF	2 ounces	Difficult to go beyond these specifications.
High-Level Mixer-Varactor	2-10	100 mw (achieved)	300 mw RF 100 mw IF	5 db maximum conversion loss RF to RF	2 ounces	The power output of these devices will increase as the state of the art advances.
RF Amplifier	2-4	2 watts at 2 GHz projected	gain 30 db	30%	<2 ounces	Design goal of MERA program.

Table 4-9. RF amplifiers and high-level mixers.

systems because intermodulation problems arise in AM systems; both division and multiplication are performed to the same factor (2, 3, etc), since division is a multi-valued function. That is, sin $(2\omega t + 2\phi)$ divided by 2 can be sin $(\omega t + \phi)$ as well as sin $(\omega t + \phi + \pi)$; but the second harmonic of either of these is the same. As will be seen from a later example, frequency division is generally performed with a conventional phase-locked loop except that the usual voltage-controlled oscillator (VCO) is replaced with a VCO followed by a harmonic generator. The output of the VCO is thus the input frequency divided by the harmonic used.

In the instance of a single frequency, such as a pilot frequency, division may be performed by first down-converting to a low IF frequency, then dividing with a digital counter stage, and finally filtering to regain a clean wave shape.

Schemes which use varactor multipliers as final power output stages do so generally because a better, compact, efficient RF power amplifier is not within the present state-of-the-art. Systems can be built now incorporating such power stages; the only other RF power stage available above several GHz is the traveling-wave tube. In the UHF region, efficient transistor amplifiers are available and multipliers are not used. Work with multipliers at the Hughes Aircraft Company has shown that the following typical multipler performance can be obtained with a single varactor diode (Motorola MV1810): one watt at 5.8 GHz can be multiplied (doubled) to one-half watt at 11.6 GHz; and 3 watts at 1.9 GHz can be multiplied (tripled) to one watt at 5.7 GHz.

The development of efficient, lightweight RF amplifiers or their equivalent such as phase-locked solid-state oscillators is a critical field of needed component research.

4.5 ANTENNAS

Spacecraft and ground-based antennas are discussed in this section as components of a link designed to fulfill the specific function of providing constant communication from a spacecraft to the earth at astronomical distances. For obvious reasons, the most attention is given the down link aspects. Particular attention is also given to the low microwave region, specifically 2.3 GHz, since a frequency in this region has advantages for an all-weather ground station and is presently in use in the NASA Deep Space Instrumentation Facility.

The requirement of a constant information rate of 10^6 bits per second with a given probability of error implies a specific system performance in terms of bandwidth and signal-to-noise ratio. When the characteristics of the available transmitter and receiver are evaluated or assumed, the required performance characteristics of the overall radiating system are determined either directly or by implication. The overall radiating system is taken to include the combination of the spacecraft and the ground or relay station antenna equipment.

Two general areas that must be investigated relative to the radiating system can be identified. The first concerns questions about the signal level or gain that is provided, ways by which it may be enhanced, and limitations that are encountered. The second area embraces questions about the contributions made to the noise of the communications link, the manner in which these are introduced, and methods by which they may be minimized. These questions are, of course, interrelated and the limitations encountered are intensely practical and economic, as well as theoretical. During the study, attention was directed to both areas of investigation with only partial answers available at this time.

The requirement of a minimum signal level forces the sum of the gains of the space and ground antennas to be of some value that can be specifically determined for a particular mission. It is important to be able to allocate the antenna gains at each end of the link according to

reasonable expectations concerning the practical designs and performance characteristics that can be accomplished in the next ten years. An optimum allocation of these gains is difficult and will require much investigation. On the basis of the plans for the 1971 Voyager mission, the sum of the gains on future missions can be estimated to be about 110 db, or a 20-db increase over the gains specified in the Voyager link for which a spacecraft transmitter of 50 watts has been postulated. It becomes necessary to determine the aperture sizes and implementation techniques that can provide these gains.

Consideration is given in the next section (4. 5. 1) to the aperture size required to provide the necessary gain. This discussion is followed in section 4. 5. 2 by consideration of some fundamental limitations on the performance of high gain antennas. Although there are many factors which limit the performance, emphasis has been placed on the degradation of maximum antenna gain caused by phase errors in the effective aperture of the antenna. The cause of such phase errors may lie in mechanical imperfections in reflector antennas, in electrical errors in a phased array, or in phase decorrelation in atmospheric propagation. The result depends only on the distribution and scale of the error but is a significant factor in the limitation of the size of very large antennas.

Because of the significance of the noise level in determination of the overall gain requirement, an entire section, 4. 5. 3, is directed to a consideration of the noise that competes with the signal and is collected and introduced at the ground end of the down link. The convention of treating the noise as resulting from an equivalent antenna temperature is followed. Since the noise level is highest when the antenna beam is directed at or near a noise source, attention is paid to the case in which the spacecraft is at or near conjunction with the sun or a planet, and particularly to the effect of the sun in the antenna sidelobes.

In section 4. 5. 4, the state-of-the-art of potentially applicable antennas is presented in relation to the problems of achieving very high gain. Consideration is given to reflector antennas, phased array

antennas, and self-steering antennas. In section 4. 5. 5, antennas for the ground-based stations are discussed, followed in section 4. 5. 6 by a discussion of antennas for spacecraft. Large phased arrays and self-phasing arrays are considered for large aperture applications, and several problem areas are presented which require further study.

4. 5. 1 Gain and Aperture Requirements

In view of the background material discussed above, it is possible to make some general assessments of the gain and associated aperture required to provide nearly continuous communication between the ground and the spacecraft of future missions. The planned characteristics of the Voyager spacecraft to be used in the 1971 Mars probe and the characteristics of the Jet Propulsion Laboratory 210-foot paraboloid afford a convenient reference, since these characteristics represent the state-of-the-art. Lacking more definite data, it is possible to use Voyager characteristics as a basis to determine the incremental changes necessary to achieve an information rate of 10^6 bits per second for, say, 1 AU of range. The data of Table 4-10 are from JPL[10] and include the Mariner IV Mars probe of 1965 for comparison. Of course, both systems operate at 2. 3 Ghz since they are part of the present plans which call for use of the DSIF stations.

For Voyager, each antenna has a 55 percent efficiency, the overall losses are 10 db (pointing error, system, tolerance, lost power in sidebands, etc.), and the probability of error per bit is $P_e = 10^{-5}$. The rate parameter, $\mathcal{R}_o$, is 1.3×10^5 per sec, and for biphase coding, $\mathcal{R}_o/\mathcal{R} \simeq 10$ and $B/\mathcal{R} \simeq 1$ (Figure 2-1). The Voyager spacecraft data are projected design data and presumably represent the forefront of the art in view of spacecraft technological problems and booster capability. It can be noted that the Voyager spacecraft, as planned, represents an improvement of approximately 15 db in the product of gain and power of the transmitter ($P_T G_T$) or the effective radiated power (ERP) as compared with Mariner IV. It can be reasonably assumed that the ERP can be further improved by straightforward designs of the spacecraft

Program	Space Transmitter		Ground System		Data Rate (bits/sec)
	P_T (watts)	G_T (db)	G_R (db)	T_a (°K)	
1965 Mariner IV 575 pounds	10	24 (3-foot dish)	53 (85-foot dish)	55	34 (planned but not achieved)
1971 Voyager 7000 pounds	50	32 (7-foot dish)	61 (210-foot dish	25	12,000
Projected					10^6

Table 4-10. JPL Communications Capability at 1 AU.

antenna and associated components and by increase of the transmitter output power as allowed by expected enhancement of propulsion and spacecraft performance.

Extrapolating from the planned Voyager characteristics, it can be observed that to increase the data rate to 10^6 bits/sec requires an overall increase in effective link gain of about 20 db. How this 20 db increase is divided between the ends of the communication link is a practical and economic matter that depends on capabilities of the booster and spacecraft components and on the communication techniques to be used. If it is assumed, for instance, that no further burden can be placed on the spacecraft components, then the ground system must provide all the gain enhancement with a resulting gain of about 80 db at 2.3 GHz. As a point of reference, this gain is equivalent to that of a circular aperture with a diameter of approximately 2000 feet. A parabolic dish of this size is relatively impractical, since it must be assumed to have the same surface tolerance and illumination efficiency as the 210-foot JPL dish and to maintain the same noise temperature and greater pointing accuracy. Arrays on the order of 2000 feet on a side are perhaps somewhat more feasible but certainly costly. Some of the required improvement can be transferred to the spacecraft components to make the size of the ground antenna more manageable.

A 10-db improvement in the gain of the antenna on the vehicle would require, for the same level of aperture efficiency as the Voyager antenna, a ten-fold increase in area. Such an area increase corresponds to an aperture diameter increased to about 22 feet with a resulting half-power beamwidth on the order of 1.2 degrees at 2.3 GHz, as opposed to roughly 3.7 degrees for the Voyager 7-foot dish.

With the 10-db gain improvement in the spacecraft antenna and some enhancement of transmitter power, it can be anticipated that the gain of the ground antenna will now lie between 60 and 80 db. The diameters of circular apertures corresponding to these gain bounds at 2.3 GHz are slightly less than 200 and 2000 feet, respectively, with the supposition that the beam is formed perpendicular to the aperture and that an allowance is made for taper and other losses. The 3-db beamwidths are on the order of 0.13 degree and 0.013 degree, respectively.

It is a reasonable working assumption that the improvements can be effected at both ends of the link. The result would be that the gain of the ground antenna would fall in the neighborhood of 70 db and its active area would correspond to that of a circular aperture of about 600 feet in diameter. The noise energy collected by the antenna affects the final determination of aperture size, however, and noise considerations are taken up in succeeding sections. The noise is important in determinations of the required sidelobe level and the necessary taper of the aperture distribution. Cases of particular interest occur when the spacecraft approaches a radiating body such as a planet or the sun, with the near approach to the sun in angle constituting a limiting case.

4.5.2 Phase-Error Limitations

Random phase errors across the aperture illumination can degrade the gain, increase the sidelobe level, and cause null-filling. The phase errors may result from a variety of causes and have vastly different correlation distances that depend on the nature of the source. In general, they can be divided into three classes:

- Small scale phase errors of a truly random nature whose correlation interval is small, much less than the dimension

of the aperture but greater than the wavelength. This class of errors is associated with manufacturing tolerances and certain atmospheric effects which tend to have small correlation distances.

- Slowly varying phase errors over the aperture whose correlation interval is on the order of the characteristic physical dimension of the aperture. This class of error is almost systematic and arises from physical distortions, such as bending or warping of the main support structure, or from any large scale distortion of an incoming wavefront.
- Phase errors associated with the individual elements or subarrays whose correlation interval is intermediate between the other two classes of errors. Examples of these intermediate errors are the positional errors in the antenna mounting to the support structure and the errors in amplitude and phase of the signals fed to each element by the feed system.

Effects Due to Random Errors. The degradation of two-dimensional antenna gain due to random phase errors has been analyzed statistically by Ruze.[11] The phase errors of concern here are those random phase errors caused by loose machining tolerances and random distortion of the aperture surface. Ruze considered both discrete array and continuous aperture antennas. In general the same statistical considerations apply to each type. In the discrete array, the error in one array element is independent of the errors in adjacent elements. However, this assumption does not hold for a continuous aperture antenna since if the error is large at one point, it will probably also be large in the immediate neighborhood. Therefore, a correlation interval, C, is defined as that distance, on average, at which the errors become essentially independent.

Ruze shows that the gain of a continuous aperture is degraded to a greater extent as the correlation interval is increased. This relationship is shown graphically for a parabolic reflector antenna in Figure 4-5 in which a family of curves representing different values of C is plotted.

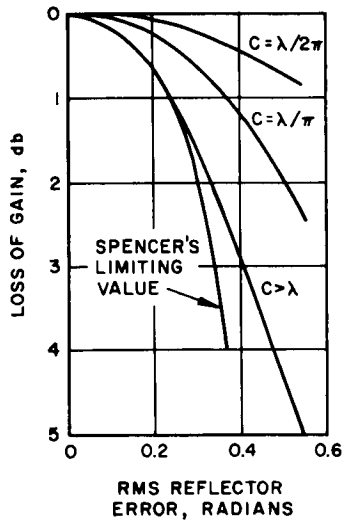


Figure 4-5. Reduction of gain of parabolic mirror (from Ruze).

It should be noted that gain drops sharply for large values of C. Errors with both large and small correlation intervals can be expected in either large reflectors or large fixed arrays; hence, the utmost care in fabrication must be exercised to keep the correlation intervals small, or at least to keep the errors associated with large correlation intervals to very small values.

Distortions of the reflecting surface of reflector antennas cause twice as much degradation of gain as do equal distortions in phased arrays because of the two-way path of a reflected signal. Figures 4-6 and 4-7 present the error-free antenna gain and gain with random phase errors as functions of aperture size for a correlation inverval of $C > \lambda$. The root mean square phase error in each figure is a particular fraction of the electrical diameter. Therefore, the rms phase error on a given curve becomes larger as the diameter increases, and the error-free gain and random phase error gain curves diverge. The effect of such errors on the sidelobe level and pointing direction of large arrays has been studied by both Elliott [12] and by Rondinelli. [13] The latter has very conveniently computed maximum sidelobe level within a specified cone and also everywhere outside such a cone, in probabilistic terms, using the well-known Q-functions of Markum. [14] Several different random variable distributions are considered.

For a given tolerance, pattern deterioration is found to decrease as the array is enlarged. In addition, for the same tolerance, pattern deterioration is less for a planar array of size L^2 than it is for a linear array of length L. Finally, the increase in sidelobe level due to random errors may be shown to be independent of scan angle for scanning antennas.

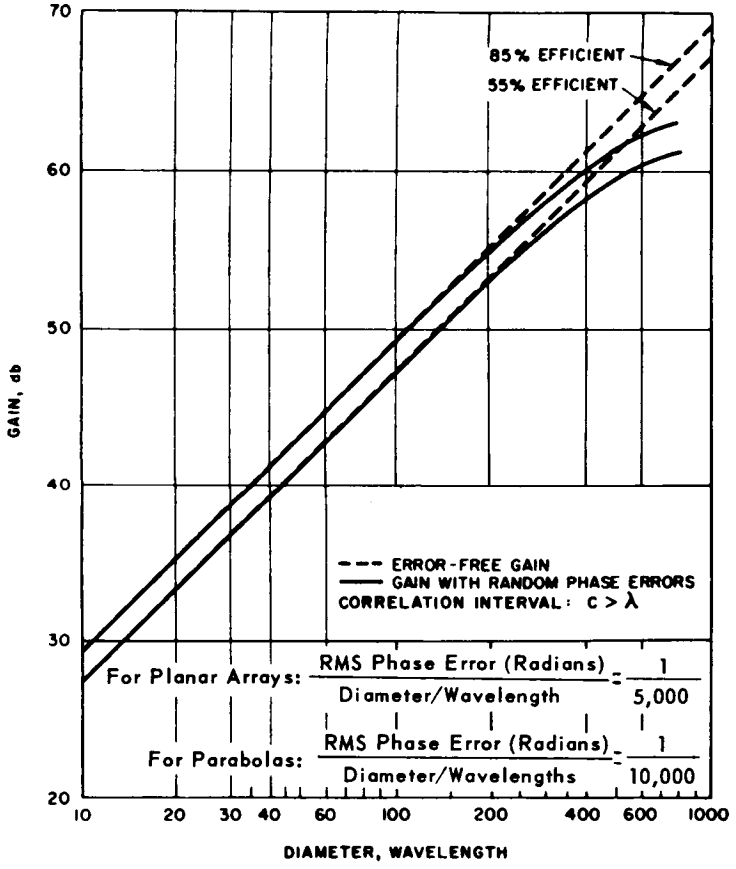


Figure 4-6. Antenna gain degradation due to random phase errors in aperture illumination.

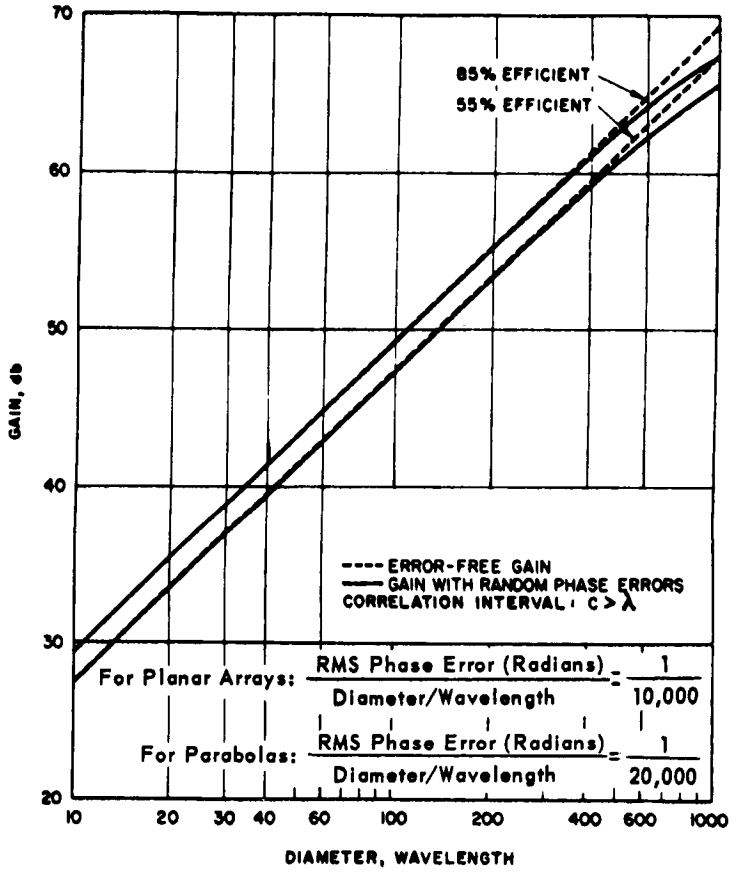


Figure 4-7. Antenna gain degradation due to random phase errors in aperture illumination.

Effects Due to Systematic Errors. Systematic phase errors (those with a large correlation interval) across the aperture of a large array of small elements, a large array of sub-apertures, or even across the continuous aperture of a large reflector antenna can result in a significant loss in gain. Bailin[15] has reported on losses of this type. Effects on gain for certain types of systematic distortion were calculated for the first three bending modes of distortion and were found to be similar. Figure 4-8 shows the degradation in gain plotted for the first and third bending modes for a paraboloid to be used in space. This type of deviation approximates the kind of distortion expected from the thermal environment on a space mission. The data shown in Figure 4-8 are based on Bailin's theoretical work[15] for uniformly illuminated linear apertures and are a conservative estimate for the effects of sinusoidal distortion in a radial mode on the gain of a circular aperture with a tapered illumination. An evaluation of sidelobe level and null filling may also be made from Bailin's equations.

The deflection of an aperture surface due to construction techniques or thermal effects tends to increase linearly with the diameter of the reflector; consequently, the size of the reflector is limited by the achievable tolerances, in addition to other limiting factors. The relationship between the reflector or aperture size and the tolerances is illustrated in Figure 4-9. Tolerances are given in terms of ratio to the diameter so that each curve may be related to a particular fabrication technique, and deviations tend to increase with the size of the reflector. The situation and the results are similar to the random error case discussed above.

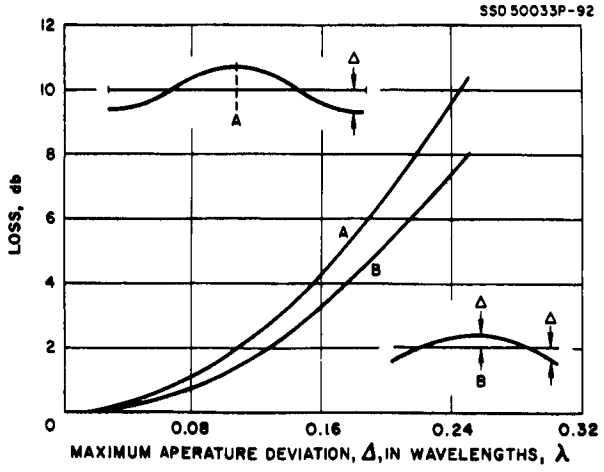


Figure 4-8. Degradation in gain due to sinusoidal deflection of paraboloid.

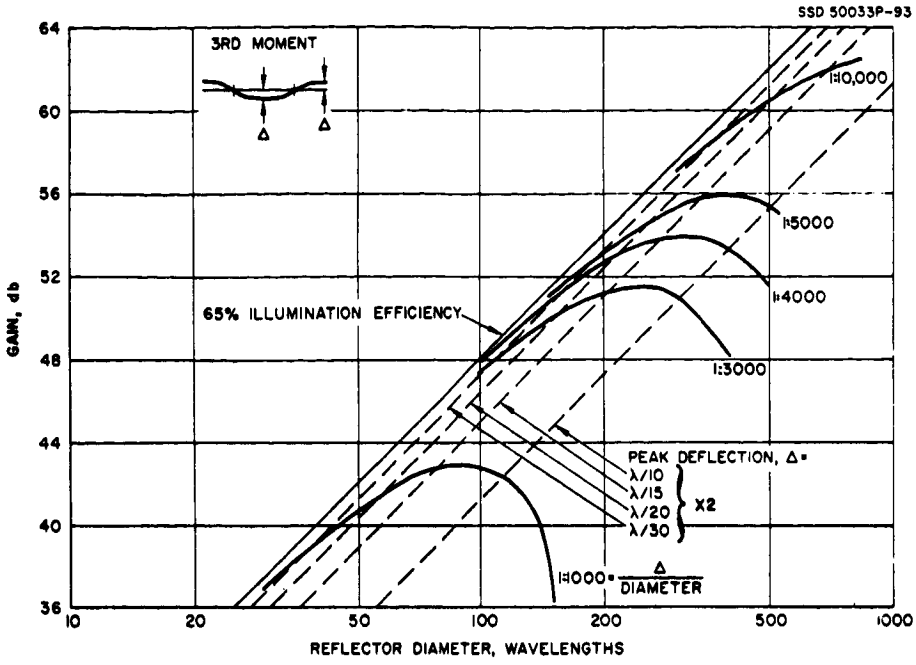


Figure 4-9. Gain versus reflector distortion.

Clusters of large apertures would also have a tendency to generate grating lobes because of gaps in the aperture and because of phase errors that may exist in the electronic equipment used to tie the individual elements together. A large phased array of smaller elements would also be subject to systematic errors of the type affecting the paraboloids even though such an antenna would have essentially no gaps in the aperture. Thus, the data presented for paraboloids may also be indicative of systematic phase errors in large arrays of a discrete element or clusters of elements in subarrays.

Pattern Degradation Due to Errors of Intermediate Size. The effects of phase errors of the third class on sidelobe structure, gain, and beam pointing direction can be conveniently and easily studied parametrically as a function of element or subarray spacing, frequency of operation, and error distribution by means of one of the several medium-sized time-shared computer systems now available to engineers. By incorporating a pattern plot routine into the computer program, a study of the

complex far-field radiation as a function of the error parameters is made. Meaningful results can be obtained in a few minutes once a relatively simple program has been established for a particular antenna configuration or set of parameters.

Atmospheric Effects and Considerations of Gain Degradation. A semi-quantitative discussion will be given in this section of the effects produced by turbulence of the atmosphere on the propagation of electromagnetic waves and on gain degradation of large aperture antenna systems. The atmosphere is in continuous turbulent motion and is, therefore, neither uniformly stratified nor homogeneously mixed. The exact structure of the turbulence is an extremely complex function of space, time, and meteorological state. If these atmospheric irregularities can be considered randomly varying, the turbulence may often be described in the form of a statistical model which may be of use in investigating a varied number of effects arising in propagation. These types of investigations have been carried out in detail by a number of authors.[16]

In order to give a description of the possible effects involved, it is usually assumed that the correlation in the refractive index at two different points of the atmosphere is only a function of the separation of the points and independent of time. The suppression of the time dependency is considered justifiable since the variations occurring in the atmosphere are relatively slow as compared with the information rate normally used in communication. With this assumption, the correlation function for the atmospheric dielectric constant can be given by the expression

$$C(r_o) = \frac{<\Delta\epsilon(r)\Delta\epsilon(r + r_o)>}{<(\Delta\epsilon)^2>} \qquad (1)$$

which describes the cross correlation between simultaneously varying values of permittivity at two locations, separated by a distance r_o. The magnitude of the phase and amplitude fluctuations of a planar wave may be obtained from integral expressions of $C(r_o)$ derived with an

appropriate model of the atmosphere considering the irregularities.[9] If ℓ_o is assumed to constitute the average size of the atmospheric irregularities or blobs, the correlation function can be conveniently stated as a function of r_o/ℓ so that the scale length or magnitude of atmospheric turbulence can be taken into consideration.

Since the magnitude of atmospheric blobs will vary, they contribute different distributions of phase errors which may be classified by their correlation intervals into the three classes of errors discussed previously. The correlation function used to describe scintillation error due to rain will have a small correlation interval since the blob sizes are small. The non-uniformities in the rainfall will, on the other hand, contribute scintillations that vary at a less rapid rate but are of great magnitude. The blobs causing these will have a correlation distance which is greater than a wavelength and, possibly, as large as the aperture size of the antenna. Phase perturbations caused by variations in dielectric constant of the atmosphere will be caused by atmospheric blobs which will also have correlation distances that are large relative to a wavelength and that often extend beyond the aperture size of the antenna. In general, the correlation distance will depend on the particular type of atmospheric irregularity with which it is associated.

For a discussion of the process required to determine the phase perturbation in the wavefront, it is assumed that the dielectric variations are uncorrelated. A ray emanating from a point source after traveling through one blob will have a mean square deviation from the mean of its phase front that is equal to $\ell^2 < (\Delta\epsilon)^2 >$ where $< (\Delta\epsilon)^2 >$ is the mean square fluctuation of the refractive index of the blob. At a distance R from the source, the ray will have propagated through R/ℓ uncorrelated blobs, and the mean square deviation from the mean of the phase will become $R\ell < (\Delta\epsilon)^2 >$. This value can be used directly to determine the gain degradation for an aperture of gain, G_o, from the results of Ruze:[11]

$$G = G_o \left[1 - < \delta^2 > \right] \qquad (2)$$

$$G = G_o \left[1 - R\ell < (\Delta\epsilon)^2 > \right]$$

where G is gain of the system operating in the turbulent medium. The evaluation of $<\delta^2>$ is complex and will not be attempted here. The estimation of the phase perturbations depends on the correlation model employed and on the position and orientation of the antenna relative to the turbulent medium. The results, therefore, can vary considerably depending on the assumptions made.

It should be emphasized that very little conclusive experimental evidence exists to substantiate the validity of any of the analyses that have been performed, especially for large aperture systems, on which atmospheric turbulence may prove to have a considerable effect for adequate system performance. Although many experiments concerned with signal scintillation studies and correlations in scintillation between two or more apertures at various spacings have been made,[17] the results give no information on actual phase variations across the apertures themselves which is required to estimate the degree of gain degradation involved. For the studies that have been performed, relatively little correlation with existing meteorological conditions was accomplished so that system performance for varied atmospheric conditions could be analyzed.

4.5.3 Antenna Noise Temperature

The characteristics of a large, ground-based antenna system for one end of a deep space link are, to a great extent, determined by the noise that competes with the potentially weak signal from a distant spacecraft. The noise received or the equivalent antenna temperature under certain special conditions of a deep space mission is especially important in determining the overall gain requirement, the required sidelobe level, the average value of the sidelobe level, and the necessary distribution taper. In an effort to establish some criteria for the selection of certain of the critical antenna design parameters, consideration has been given to the noise temperature contributed by losses in the antenna itself and by atmospheric attenuation, by the presence of a warm earth, by nearness to planets, by galactic noise, and by the sun in the antenna sidelobes.

Considerable information on extraterrestrial noise and attenuation has been published.[18-23] Most of the available results have been summarized in two publications.[10, 21] One of the most important antenna characteristics of a high-gain system is the response to various warm objects in the sky and on the earth. In a reflector antenna, both the main reflector pattern and the spillover from the primary feed pattern contribute to the noise temperature. For an evaluation of the antenna temperature, the entire diffraction pattern of the antenna should be considered. The basic equation required to calculate the antenna noise temperature is

$$T_a = \frac{\int T(\theta, \phi)\, G(\theta, \phi)\, d\Omega}{\int G(\theta, \phi)\, d\Omega} \tag{3}$$

where

$T(\theta, \phi)$ = noise temperature as a function of an angular position in space

$G(\theta, \phi)$ = antenna gain as a function of an angular position in space.

$d\Omega$ = solid angle element

The numerator of Equation (3) may be used to determine the effects of various warm objects as they interact with the different portions of the antenna pattern. Thus, by a subdivision of the above integral, Equation (3) may be written as

$$T_a = \frac{\left[\int_{\Omega_B} T_E\, G_B\, d\Omega + \int_{\Omega_M} T_S\, G_M\, d\Omega + \int_{\Omega_{B'}} T_S\, G_B\, d\Omega \right]}{\left[\int G\, d\Omega \right]} \tag{4}$$

where

T_E = earth temperature

T_S = sky temperature

G_M = main lobe gain

G_B = minor lobe gain

Ω_M = space integral over main beam

Ω_B = space integral over the minor lobe with an earth background

Ω'_B = space integral over the minor lobe with a sky background

When G is properly normalized, the denominator becomes

$$\int_{\text{all space}} G \, d\Omega = 4\pi$$

or slightly less because of ohmic losses in the antenna. The effects of warm objects on antenna performance can be estimated from an examination of the contribution of each of the portions of Equation (4).

Antenna Loss. One important source of noise which is not covered by Equation (4) but nonetheless competes with the signal arises from the ohmic losses in the antenna itself. If the sky temperature is T_S, the antenna has a power transmission coefficient τ, and the antenna is at ambient temperature T_o, then the effective antenna noise temperature is

$$T_a = \tau T_S + (1-\tau) T_o \qquad (5)$$

Thus, 1 db of loss contributes 60°K if T_o = 300°K. Losses of this level are common in present-day phased arrays because of inefficient phase shifters and the aperture mismatch problem in matched-feed phased arrays. These contributions to noise can be reduced only by improved component design, by the use of low-noise amplifiers, or by cooling the entire array if it does not impair any components. Although considerations of this kind concerning the various components and their interconnections are a part of good antenna design, the availability of the actual components in a suitable form will greatly influence the

selection and antenna type for any large aperture. Any attenuation effects in the atmosphere can be treated like antenna loss since they are in series with respect to the overall system performance.

Warm Earth. An important noise contribution is made by warm objects from which energy leaks into the minor lobes of the antenna pattern. The noise leakage from the warm earth, even though it is largely excluded by the antenna directivity, is considerable because it enters from almost all the back lobe hemisphere. It can be seen that the first and third integrals in Equation (4) cover the space occupied by the minor lobes and are to be taken over the lower and upper hemispheres except for the main beam angle. This exception is very small for a large aperture. The main beam contribution is represented by the second integral. If it is assumed that G_B, T_E, and T_S are fairly constant over each of the appropriate hemispheres or that they represent the average values, then the equation for T_a without the main lobe contribution becomes

$$T_a = \frac{G_B}{4\pi} \left[T_E \int_{\substack{\text{lower} \\ \text{hemisphere}}} d\Omega + T_S \int_{\substack{\text{upper} \\ \text{hemisphere}}} d\Omega \right] \tag{6}$$

Since the integrals are essentially taken over hemispheres for which

$$\int d\Omega = 2\pi$$

then

$$T_a = \frac{G_B}{2} (T_E + T_S) \tag{7}$$

This equation shows the dependence of the antenna temperature on the average minor lobe gain, G_B. The factor of 1/2 merely states that essentially one-half the space surrounding the antenna has an earth background while the other half consists of a sky background. Thus the effect of the warm earth may be calculated from the backlobe radiation characteristics.

The contribution to antenna temperature from the warm earth must be kept to a minimum in an extremely low-noise receiving system (system noise temperature of 10° K to 40° K). The required range of average minor lobe gain may then be determined. For instance, if the maximum allowable contribution to the antenna temperature due to the warm earth is taken as 5° K, then the average minor lobe gain must be 14.9 db below isotropic; similarly, 10° K corresponds to an average minor lobe gain of 11.9 db, and 15° K* corresponds to 9.7 db below isotropic. Stated in a slightly different way, the average minor lobes must be 84.9 db, 81.6 db, and 79.7 db below the main lobe of a 70-db antenna in order to maintain a contribution to antenna temperature due to the warm earth of less than 5° K, 10° K, and 15° K, respectively.

Sky Noise as a Function of Frequency and Weather Conditions. A distribution of galactic noise can be considered next. The antenna temperature T_a is related to the source distribution T_s and the antenna gain G by

$$T_a = \frac{1}{4\pi} \int G\, T_s \, d\Omega \simeq T_s \qquad (8)$$

where the last approximation is valid for a hot region that is larger than the beam. Distributed noise is most serious along the galactic equator, particularly in the direction of the galactic nucleus near the constellation Sagittarius. Unfortunately, the ecliptic passes through Sagittarius and cuts the galactic equator again near the constellation Taurus. Galactic noise has a temperature dependence as $f^{-5/2}$; consequently, frequencies above 1 GHz are favored as the operating band for a deep space communication system. On the other hand, the atmospheric absorption and the atmospheric noise in bad weather are serious, and frequencies below 7 GHz are indicated. JPL's DSIF system has settled on 2.3 GHz, and the noise temperature for their 210-foot reflectors is expected to be 25° K. [10]

*An antenna temperature of 15° K has been reported for an 85-foot reflector measured at 2.388 GHz. [22]

This result disregards the effect of discrete noise sources which contribute to the antenna temperature in accordance with the flux density H (in watts per meter^2 per Hz in the polarization of the antenna), the effective aperture A_a, and Boltzmann's constant k:

$$T_a = \frac{HA_a}{k} = T_s \left(\frac{\theta_s}{\theta_a}\right)^2 \qquad (9)$$

where θ_s and θ_a are the angles subtended by the source and the antenna beamwidth respectively. Fortunately, most of the radio stars are not in the zodiac and may be disregarded insofar as the contribution to the main beam is concerned. At 2.3 GHz the most intense radio star in the zodiac is Taurus A with a flux of about 8×10^{-24} watts per sq. meter^2 per cps and a frequency dependence of about 1/f.[23] The remaining sources are the planets, the moon, and the sun which are discussed in the next two sections.

The sun is by far the most significant source with a temperature of 10^5° K at 2.3 GHz and a 1/f dependence. When the antenna is looking at the planets near conjunction, the sun will be a major source of interference because of its large 1/2-degree disc and high temperature. Since the sun and some of the other discrete sources have 1/f frequency dependence, the 2.3 GHz used by JPL is not optimum for discrimination against these sources. For example, when the range is extended from 1 to 10 GHz, the sun's temperature drops by a factor of 10, but the atmospheric attenuation is small and changes little over this band. Where the antenna may be looking close to the sun, therefore, a frequency closer to 10 GHz would be better. The optimum frequency is clearly dependent on mission requirements; 2.3 GHz is a good choice in terms of galactic background noise and rainy weather conditions, but a higher frequency will have advantages for operation against discrete noise sources. More study is required to select the single optimum frequency or to select different frequencies for different conditions. It may be noted that during the 1971 and 1973 Martian optimum opportunities, Mars is viewed at night from each station and the sun

plays no role. Similarly, when the antenna is looking at the inferior planets near greatest elongation, the sun plays a negligible role. On the other hand, for missions of long duration it may be necessary for the station antenna to look at a vehicle in close angular proximity to the sun. The effects of strong discrete sources in the sidelobes are discussed in following sections.

Noise Temperature Due to Planets. Once the order of magnitude of the aperture size has been established, it is important to examine the values of the sky temperature as seen by apertures with narrow mainbeams and low sidelobes that are achieved by tapering the illumination functions. The sky temperature seen by an antenna with beamwidth θ_a when pointed at a hot planet of temperature T_p, and disc θ_p (angle subtended by the planet at the earth), is obtained from the second term in Equation (4) as

$$T_a = T_p \left(\frac{\theta_p}{\theta_a} \right)^2 \qquad (10)$$

if $\theta_p \leq \theta_a$. If the object fills or is larger than the beam, $T_a = T_p$. For circular apertures with illumination functions of the form $(1 - r^2)^2$ which corresponds to an illumination efficiency of 56 percent, the peak sidelobe is approximately 31 db. The 3-db beamwidth of a 2100-foot aperture is 62 seconds of arc and that of a 1000-foot aperture is 130 seconds of arc (at 2.3 GHz). These beamwidths are on the order of, or larger than, any planetary discs except the moon and sun whose temperatures are 200° K and 10^5° K respectively at 2.3 GHz. The results for the nearer planets calculated for two antenna sizes and on the basis of planet temperatures as given in Reference 12, Section 2.6, are listed in Table 4-11.

From Table 4-11, two items are apparent. First, the temperature seen by the mainbeam of an antenna when looking at a hot planet increases with increasing size of the aperture if $\theta_p \leq \theta_a$. Second, when the beam points directly at a hot source, the noise temperature can exceed a nominal design level of 25° K and, consequently, the

Planet	Source Temperature °K	Maximum Disc (seconds)	Noise Temperature °K	
			2100-foot reflector	1000-foot reflector
Venus	580	65	580	145
Jupiter	800	48	480	110
Mars	217	25	35	8

Table 4-11. Effective noise temperature at 2.3 GHz due to planets for two antenna apertures.

information rate and bandwidth must be reduced by large factors to maintain the error rate. It is also to be noted that an increase in the aperture size does not improve these factors until the beamwidth is smaller than the planetary disc. This is apparent because the information rate $\mathcal{R}$ depends on the parameter $\mathcal{R}_o$ which is a measure of the ratio S/N. For $\theta_p \leq \theta_a$ and $T_a = T_s$ (no environmental noise or losses in the antenna),

$$\mathcal{R}_o \sim \frac{G_R}{T_s} \sim \frac{\frac{1}{\theta_a^2}}{T_p \left(\frac{\theta_p}{\theta_a}\right)^2} = \frac{1}{T_p \left(\theta_p\right)^2} \tag{11}$$

which is independent of the aperture. However, for $\theta_a < \theta_p$, $T_s = T_p$ and $\mathcal{R}_o$ is proportional to G_R, the gain of the receiving antenna, which varies with the area. Thus, by using a larger aperture, the information rate $\mathcal{R}$ could be increased for communication with probes in conjunction with Venus and maybe even Jupiter, but very large increases over 2100 feet in diameter would be required for improvement in the cases of the other planets. Alternatively, a small aperture will do as well as a big one when the space vehicle is in conjunction with the large hot planets. In any case the data rate will be low since it is limited by $\mathcal{R}_o/B$.

Effects of the Sun in the Sidelobes. The most intense of the discrete astronomical noise sources as seen from the earth is, of course, the sun, which can introduce a significant noise contribution through the antenna sidelobes. The sun is a particularly effective contributor to the noise level when communications are to be maintained with probes on missions to the inferior planets. Since it is apparent that when it is in the mainbeam of the antenna, the sun with a noise temperature of 10^5° K (at 2.3 GHz) would saturate any presently conceivable receiving system, it is imperative that a study be made of the effects produced by the presence of the sun in the sidelobe region of the radiation pattern. An analytical technique has been developed for evaluating these effects quantitatively, at least under worst conditions in which the antenna temperature is evaluated along the envelope of the far out side-lobes. The principal objective of this analysis was to determine the optimum distribution of the excitation across the aperture as it relates to total cost of the ground antenna and the number of days that noise from the sun precludes high data rate communication with the spacecraft.

In Appendix A, a calculation is made to determine an angle θ_o (shown in Figure 4-10) which represents the closest approach that the peak of the main beam may make to the sun without having the antenna noise temperature exceed an arbitrary value of 25° K. The angle θ_o is then reduced to a parameter ρ which can be described as an effective diameter of the sun normalized to the optical diameter. The peak of the main beam of the ground antenna must then avoid this effective diameter of the sun to assure continuous communications. This technique may, of course, be used to calculate the effects of other large noise sources in the sidelobes of the ground antenna.

An attempt is then made to estimate the occultation time, N, for a typical Jupiter trip as a function of ρ and the illumination function of a circular aperture of the form $(1 - r^2)^p$ where p is a small integer. * The illumination is characterized by the level of the peak of the first

*This function was chosen for convenience because the radiation characteristics are available in Silver. [24]

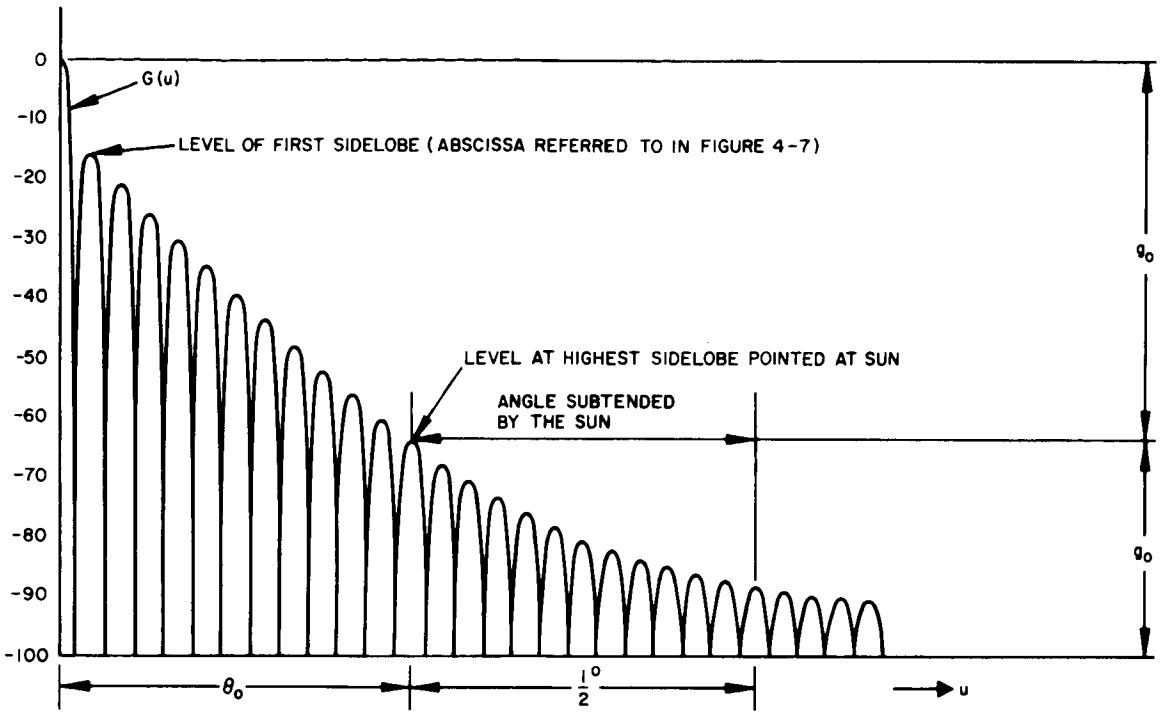


Figure 4-10. Circular aperture pattern.

sidelobe level, because this is the convention in antenna design. The occultation time is defined, for present purposes, as the time during which the sun's contribution to the noise temperature of the antenna equals or exceeds 25° K. A determination is made of the size of the aperture required to maintain a given peak gain as the distribution is altered to lower the sidelobe level. The size thus determined is incorporated in the expression for ρ in such a way that ρ and N can be plotted versus the first sidelobe level for a given peak gain. As shown in Figure 4-11, ρ drops very quickly with the first sidelobe level for various values of taper and levels out at about 25 db. This same effect is shown more dramatically in Figure 4-12 in which N is shown plotted against the height of the first sidelobe which determines the amount of taper. It can be seen that it is not fruitful to consider antenna designs with very severe taper, since it will not improve the ρ or the occultation time very much.

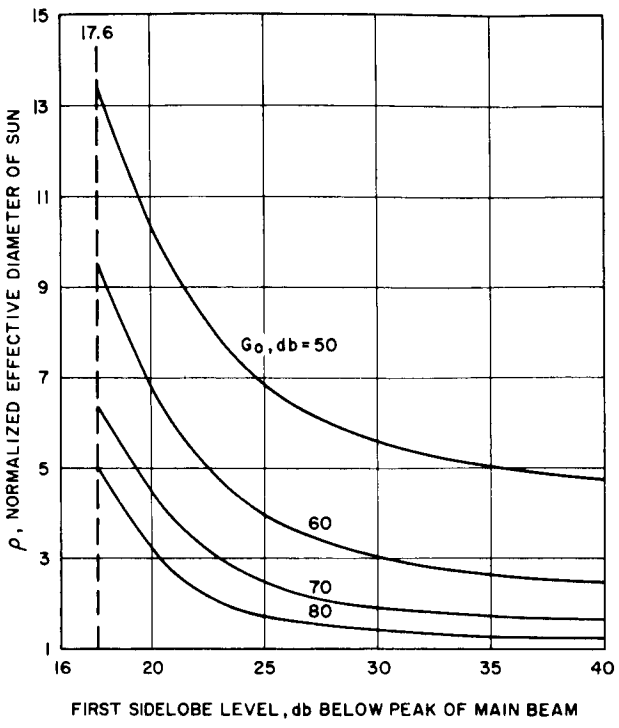


Figure 4-11. ρ versus first sidelobe level for several values of G_o.

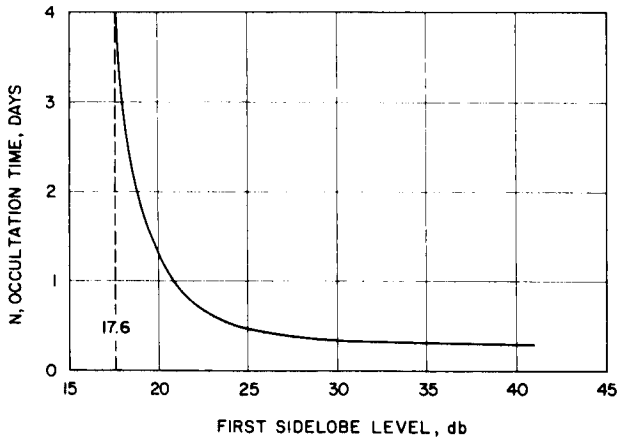


Figure 4-12. Lost communication time.

An important aspect of the design of a ground-based antenna for space communication is the trade-off between antenna cost (not including real estate) and the capability of pointing the antenna beam near the sun. This problem is examined in detail in Appendix B in which the cost of a single ground station that yields 70 or 80 db gain is estimated. A circular array of parabolic reflector antennas is used for this purpose since costs for such antennas are well established. It is assumed that the antenna systems consist of arrays of 85-foot and 210-foot paraboloids and that the illumination function is the same as is used in Appendix A. Then for a fixed peak gain, the size of the aperture, or the number of parabolic "elements", required will be a direct function of aperture efficiency and indirectly of the first sidelobe level. The expected cost thus becomes a function of the first sidelobe level and a curve of this relationship is plotted in Figure 4-13 on

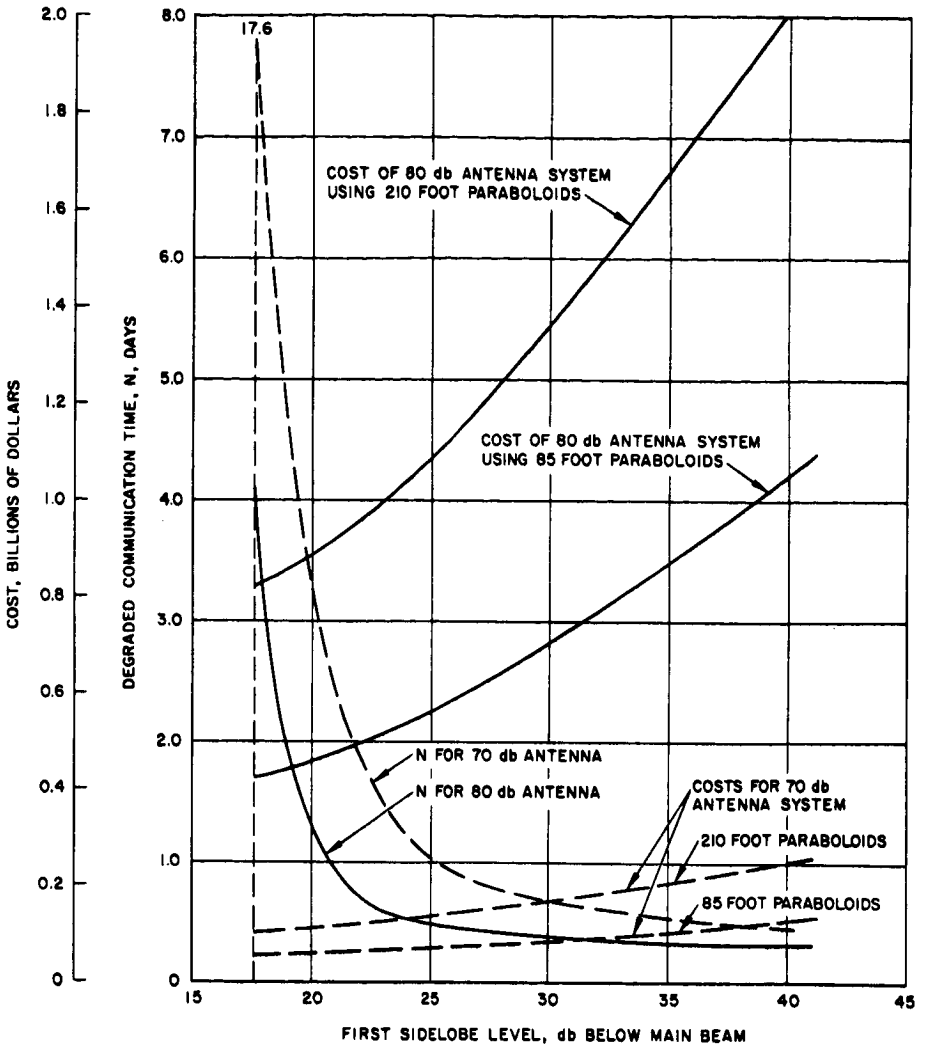


Figure 4-13. Occultation time versus first sidelobe level.

the same graph with N, the occultation time, versus first sidelobe level. Then the cost of reducing N by using a tapered distribution to reduce the sidelobe level while maintaining a fixed peak gain is easily visualized.

The conclusion to be drawn from the analysis is that antenna systems of high aperture efficiency (uniform illumination) cannot be pointed near the sun without degrading system performance. However,

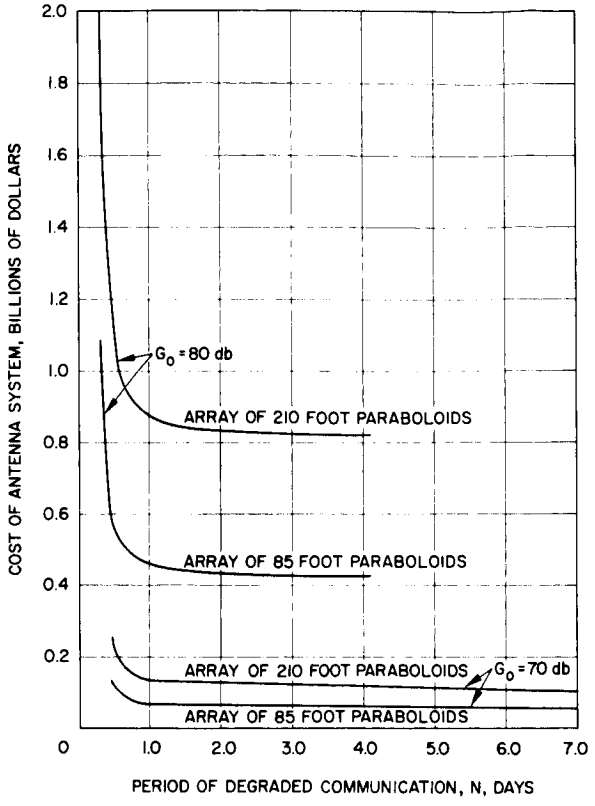


Figure 4-14. Estimated costs of antenna systems versus occultation time.

considerable improvement can be effected if some taper in the illumination is employed and if the operational requirements are sufficiently important to justify the added cost and the decreased efficiency. Figure 4-14 shows the estimated costs of antenna systems with constant gain of 70 and 80 db versus the occultation time. It can be seen that for an 80-db antenna, the occultation time can be cut from 4 days to 1 day by spending about 10 percent more than the cost of the arrays with uniform illumination. This cost increase corresponds to designing for first sidelobes which are approximately 21 db below the mainbeam instead of 17.6 db. For a 70-db antenna, cutting the value of N from 8 days to 1 day requires a tapered illumination with a 25-db first sidelobe. The results show that reducing the occultation time in this manner to a value of less than about one day is very expensive and suggests that either N be allowed to remain on the order of one-half to one day or that other ways should be

investigated for reducing it. One method that might be worth considering to avoid occultation during flight is to deliberately fire the probes into a trajectory that is slightly out of the ecliptic so that conjunction between the effective diameter of the sun and the probe is avoided.

Because of the significance of the sun as a noise source and because of the interesting aspects of self-steering antennas, an analysis was conducted of the effect of the presence of the sun in the pattern of the subaperture of a large adaptive array. This analysis and the conclusions drawn are included in Section 4. 5. 5 (Ground-Based Antennas).

4. 5. 4 State of the Antenna Art

Existing and proposed types of antennas that can be considered for use at the two ends of a deep space communication link are described in this subsection. A presentation is made of the state-of-the-art characteristics and performance of several existing antennas and of a few proposed types. Some of these antennas may have application on the spacecraft, and others may be more suitable for the requirements for the ground-based system. Of course, some of the same antenna technology can be applied at either end with only a modification in size and mechanical design. It is the intent to delineate the various antenna types that seem to have some possible application to long range communication systems whether these antennas have been developed, are in experimental laboratory form, or are only in the planning stages. Specific documentation for some of the characteristics stated is thus vague, since the devices are still being developed, and their expected ultimate performance is mentioned only for comparison.

Reflector Antennas. Reflector antennas can be designed in two basic forms: a front-fed system consisting of a relatively small feed and a comparatively large reflecting surface (most often paraboloidal) and a Cassegrain system consisting of a feed, a small convex or concave secondary reflector, and a much larger concave primary reflector similar to that of the front-fed system. When maximum antenna gain is required, the reflector size is chosen to be as large as practical and the feed is normally designed to illuminate the reflector with an intensity at the reflector edges that is approximately 10 db below that at the center. This tapered illumination reduces the sidelobes below those obtained from a uniformly illuminated aperture and reduces the energy spilled over the edges. The radiation efficiencies* of these reflector antennas are typically between 55 percent and 70 percent depending on their size. Their gain is directly proportional to their area and to the square of their operating frequency until it is limited by phase errors across the aperture. Phase errors may be caused either by a distortion of an incoming wavefront or by physical perturbations in the antenna structure in the form of mechanical deflections or manufacturing tolerances that create deviations in the reflecting surface or surfaces. The effects of phase errors are discussed in Section 4.5.2.

Reflector antennas are being employed in various sizes on space communication programs and on experimental research programs such as radio astronomy and propagation studies which use the antenna in a similar manner. The Deep Space Instrumentation Facility (DSIF) is presently equipped at five stations with 85-foot parabolic reflector antennas[25] having gains of 53 db at 2.3 GHz. A system noise temperature of 55°K is provided at each station. A 210-foot paraboloid[25] is in operation at the Goldstone, California, GSIF station, and two additional antennas of this size are planned for other stations. The

*The radiation efficiency of an antenna is here taken as the ratio (expressed as a percentage) of the on-axis gain to the gain that would be provided by a lossless, uniformly illuminated antenna of the same aperture size and shape.

expected performance of these 210-foot antennas is presented in Table 4-12. The size and frequency of operation of these antennas was determined on the basis of an economic evaluation using a level of 10 to 100 watts of available transmitter power in the spacecraft and a booster of the Saturn class of vehicles.

For space propagation research in the millimeter-wave region, the Aerospace Corporation is using a Cassegrain antenna system.[26] This reflector antenna consists of a linearly polarized feed, a primary

Azimuth coverage, deg.	±300 (from SE at Goldstone)
Elevation coverage, deg.	5 to 88 (tracking sidereal target) 4.5 to 90.5 (final limits)
Pointing accuracy, deg	0.02 pointing 0.01 tracking
Maximum angular rate azimuth, deg/sec	0.5 (wind ≤ 30 mph)
Maximum angular rate elevation, deg/sec	0.5 (wind ≤ 30 mph)
Maximum acceleration azimuth, deg/sec	0.2 (wind ≤ 30 mph)
Maximum acceleration elevation, deg/sec	0.2 (wind ≤ 30 mph)
Servo bandwidth adjustment, hz.	0.01 to 0.2
Gain at 2300 MHz, db	61
Beamwidth at 2300 MHz, deg	≈0.1
System temperature, * ^{o}K	23-25
Antenna temperature, ^{o}K	≈10
Reflector diameter, ft	210
Reflector f/D ratio	0.4235

*Includes maser amplifier, receiver, transmission line, listening feed, and the antenna pointing at a quiet sky.

Table 4-12. Expected performance of 210-foot DSIF altazimuth antenna.

parabolic reflector with a 15-foot diameter, and a hyperbolic subreflector. The expected value of the gain at 94 GHz is 70.5 db which includes effects of aperture illumination, spillover, blockage, and reflector contour deviations. The antenna is equatorially mounted and can operate up to 300 GHz, but initial operation has been restricted to 94 GHz. A Cassegrain feed system is used primarily because, with the effective focal point placed near the vertex of the parabola, waveguide losses are minimized and maintenance and accessibility of the RF equipment in the feed is simpler. Furthermore, the Cassegrain feed system provides a lower antenna noise temperature and greater flexibility since the effective focal point can be relocated by changing the shape of the subreflector. The radiation from the antenna deviates from that of a circular aperture because of the blocking effect of the support for the subreflector and the subreflector itself. This blockage has the effect of reducing the antenna gain, increasing the beamwidth, and increasing sidelobe level.[27] The computed blocked power,[28] taking into account a -13.2 db aperture taper, is 1.4 percent by the subreflector and 5.39 percent by the effects of the spars and their shadow region. The 94-GHz characteristics of the antenna are presented in Table 4-13.

Reflector diameter	15 ft (4.57 m)
f/D	0.3
Magnification factor	14.74
Beamwidth (half-power)	2.8 arc-min at 70-deg elevation angle.
	3.1 arc-min at 15-deg elevation angle (broadening caused by atmospheric turbulence)
Gain	70.33 db ±0.44 db (1σ), measured at 1.5-deg elevation angle
Aperture efficiency	53.6 percent

Table 4-13. Performance of 15-foot Aerospace Corporation Cassegrain Parabolic Reflector Antenna at 94 GHz.

In general, spillover, backscattering, and aperture blocking due to the presence of the feed or subreflector are considerations that will have to be investigated thoroughly when reflector antennas are considered for the high-gain end of the deep space communication system. For a large single reflector, these effects contribute to the noise temperature of the antenna, since the radiation from the warm earth couples to the back lobes of the antenna pattern. The study already cited[25] indicates, however, that the economics of the combination of spacecraft technology, booster capability, and ground-station size require the consideration of reflector antennas for the next 10 to 15 years of space exploration. These results could be modified by certain specified technological changes that may come about during this period.

Present space antennas for deep space communication are typified by the antenna system of the Mariner IV spacecraft. The salient feature of this system was that the spacecraft had to be attitude-stabilized when using its high-gain antenna. The high-gain antenna provided coverage from approximately 90 days after launch until approximately 20 days after encounter with Mars. The antenna was a 46.0 by 21.2 inch parabolic reflector that was illuminated by a pair of turnstile elements. The elements were arranged so that a right-hand circularly polarized beam was projected with a maximum gain of 23.5 db (at the 2.298 GHz transmitting frequency) and half-power beam-widths of 13.5° by 7.5°. When Mars was encountered, the earth was approximately 160 million miles away from Mariner IV and subtended an angle of only 1 second. The beamwidth of the spacecraft transmitting antenna could be reduced considerably to attain the necessary increase in antenna gain and still completely cover the earth with its beam. Tighter attitude stabilization would be required, however, to keep the narrower beam directed at earth.

The gain of the Mariner IV spacecraft antenna could be increased with an increase of aperture size. The most evident practical limitation on the size of spacecraft antennas is the size allowable for the shroud of a particular launching missile. However, inflatable, unfurlable, or modularly assembled and electrostatically erected spacecraft

antennas are possible practical means of circumventing such limitations. Within the present state-of-the-art as reported by several private communications from organizations engaged in the development of stowable antenna techniques, a 30 foot to 50 foot diameter reflector antenna could be assembled in space from modules to operate at 2 GHz to 3 GHz. It is expected that a 50 percent radiation efficiency could be achieved, with 45 db to 50 db of antenna gain.

Phased Array Antennas. A phased array antenna consists of an array of radiating elements with either fixed or variable relative phase differences. Those with fixed relative phase differences, such as the antennas on the Surveyor spacecraft or on advanced fire control and missile systems, require mechanical pointing. Those with variable relative phase differences, usually referred to as scanning arrays, require external controls to properly phase the elements to form a beam in a desired direction. The controls may be mechanical, electronic, or a combination of these for either the ground-based or the spacecraft antennas. However, present technology limits externally controlled scanning to ground-based antenna systems because of the obvious need for high reliability in the control system. For spacecraft antenna systems, the self-steering arrays discussed next seem to offer promise for the future.

When maximum antenna gain is required, all the elements of a scanning array are excited by an equal amount, and the relative phase between elements is adjusted for a beam normal to the plane of the array. Efficiencies of phased arrays range from 70 percent to 85 percent excluding losses of the phase shifters. However, as the beam is scanned away from its broadside position through an angle θ, the gain diminishes as $\cos\theta$ because of the decrease in aperture size projected in the direction of the beam. In addition, the impedance at the input terminals of the antenna changes as a function of the scanning angle and the gain is affected. The degree of mismatch with scan angle is determined from the shape of the pattern of each element in the array and the

realizable gain of the antenna at broadside. The element pattern is the pattern generated by a single element under the influence of mutual coupling between the elements in the array and is, in turn, a function of the element spacing and the environment of the element. The gain behavior with scanning is treated in some detail in the theoretical discussion of Appendix C. The results given there are particularly applicable to very large ground-based antennas.

It has been shown[29] that matching the elements in the aperture over a given scan range requires that the shape of the element pattern must follow a cosine dependence with scan angle. One condition necessary to satisfy this dependence is that the inter-element spacing be bounded to prevent the formation of grating lobes over the given scan range. The presence of such a lobe at a given scan angle will produce a dip in the element pattern in that direction. The magnitude of the dip can be shown to represent the amount of aperture mismatch of the elements at the angle at which the dip occurred. To prevent a grating lobe for a maximum scan angle at 70 degrees, the spacing along each side of a square lattice of elements should be approximately 0.51λ.

In general, it is difficult to achieve a cosine element pattern over a wide scan angle such as ±70 degrees. However, analytical and experimental studies made at Hughes Aircraft Company show that the elements could be matched to a maximum VSWR of 2.1 to 1 over a ±60 degree scan angle. This mismatch represents a drop-off of 0.6 db from the ideal cosine element pattern at 60 degrees from broadside.

Self-Steering Arrays. Earlier paragraphs describe some of the more conventional methods of forming narrow beams. When these conventional antenna techniques are utilized in a high-gain communication system, certain limitations and difficulties exist. Several methods for overcoming these will be given in this discussion of self-steering arrays. Although these array techniques may be applied to both ends of a long range link, they may be especially useful on board the space vehicle or when a communication satellite is used as a relay between the vehicle and the earth station since only a very limited information bandwidth is necessary for accurate beam pointing.

On a space vehicle, there is a size limitation that will depend, to some extent, on the type of antenna structure considered. Although erectable and unfurlable antennas may present an aperture that is considerably larger than the cross-section of the launch vehicle, tolerances on phasing may be a problem when large diameter-to-wavelength apertures are used and when conventional feeding and phasing techniques are employed. Then, the use of high-gain antennas in space and on earth also introduces a problem of controlling the pointing of the beam. In conventional antenna systems, pointing may be done either mechanically or electronically, but a priori pointing information may be required or some electronic tracking technique (e. g., monopulse) would be necessary to maintain the correct direction. The conventional electronic beam-steering techniques usually considered require electronically controlled variable phase shifters with their associated control electronics which may become quite complex. Therefore, alternative approaches to the realization of steerable high-gain antennas and the concomitant increase in data transmission rates are desirable. In principle, self-steering antennas automatically adjust the phase across the aperture to produce a collimated beam. Several recently developed self-steering antenna techniques will be described that may apply to deep space probes and that may allow increased data rates and minimize the problems associated with beam pointing accuracy. The problem of pointing is important on a long range system because of the large time constant inherent in any pointing correction that is applied at one end of the link but dependent upon information coming from the other end.

The class of antenna arrays known as self-steering arrays may conveniently be divided into three rather broad categories:[30]

- Switched multiple beam antennas (transdirective arrays).
- Arrays that are steered with phase inversion by mixing (self-phasing arrays).
- Arrays that are steered by the use of phase-locked loops (adaptive arrays).

These three categories are discussed in the following sections in an effort to delineate the basic principles of their operation and to give details of the various systems that have been implemented as breadboard hardware. In the following discussion the latter two categories are grouped together since they have certain similarities.

Switched Multiple-Beam Antennas. Switched multiple-beam antennas utilized appropriate switching and control circuits to select the proper beam on command from a transmitter station or automatically when a pilot signal is received.[31, 32] In the latter mode of operation, the selection is made through the use of logic circuitry that selects the antenna beam output with the strongest pilot signal and connects it to a receiver. The signal is then amplified and may be retransmitted in a second direction. This second direction is selected by logic circuitry that compares the strength of a pilot signal from that direction as it appears at the various antenna beam outputs. The transmitter is then connected to the beam output that results in transmission in the desired direction. The implementation of the technique requires fairly complicated logic and switching circuitry to connect the outputs corresponding to the proper beams to the receiver or to the transmitter. If continuous beam steering is required, continuously variable power dividers and additional complicated logic and control circuitry are necessary to accomplish the smooth transition from one beam to the next. The basic characteristics of a breadboard model of a self-steered, multiple-beam array constructed by the Antenna Department of the Hughes Aircraft Company, were included in Appendix D of the Interim Report of this study and have since been published.[33] The model was specifically designed for an orbiting vehicle. The array was built to provide communications with a station located anywhere within a 50° conical region. The coverage of this region was obtained using 16 narrow overlapping beams. The retransmitted signal was continuously transferred from beam to beam as the relative orientations of the spacecraft and station changed. The control logic circuitry proved to be quite complicated with over 1000 transistors used.

Several multiple beam antenna configurations are possible. These include multiple feed reflectors, various types of lenses, and beam-forming matrices.[30, 32] Reflectors with multiple feeds generally suffer from spillover or aperture blockage effects. Thin lenses may also suffer spillover problems unless rather complex feeding techniques are employed to shape the primary feed patterns. Many reflectors and lenses also suffer beam deterioration when scanned off-axis. While spherical or cylindrical Luneberg lenses do not exhibit either spillover effects or beam degradation with scan angle, systems that utilize Luneberg lenses require a large volume since their depth (from feed to aperture) must be essentially equal to or greater than the aperture itself. Planar arrays using beam-forming matrices overcome the latter problem to a great extent.

Self-Phased and Adaptive Arrays. Self-phased and adaptive arrays constitute antenna systems that use the incident RF energy to phase the elements so that a beam is formed in the direction from which the energy is received. These arrays are also called self-focusing and self-phasing antennas. The arrays may be contrasted with the usual electronically steerable arrays that require external sensors and information to do the steering. Here, no external commands are necessary to adjust the illumination across the aperture since, in principle, the self-steering array automatically steers the beam in the desired direction. Thus, the problems associated with pointing a narrow beam in a specified direction appear to be circumvented. In addition, self-phased arrays can compensate for atmospheric scintillation effects which may cause a loss of array gain due to lack of correlation among the signals at the various elements. An important limitation is set on the minimum size of each element in an adaptive array of elements: each element must be large enough so that in conjunction with its receiver, it will be able to detect and phase lock to the incoming signal. Once this is achieved, the several elements in the array may be locked together to realize the gain of the entire array.

Self-phased arrays may take on a variety of forms, depending on the type of circuitry used to implement them and on the sophistication of their operation. In the simplest form these arrays redirect incident energy back in the direction from which it came. These are termed retrodirective arrays. In the process, the signal is amplified, and information stored at the array is impressed on the signal before retransmission. A number of such systems are described by Kummer and Villeneuve.[30] Perhaps the simplest system of the retrodirective type is the Van Atta array.[34] In this array opposite elements of a planar array are interconnected through equal length lines to achieve the retrodirective function. Figure 4-15 illustrates the interconnections for a linear array. This type of array may be active or passive. By inserting amplification in the transmission lines between pairs of elements, the system becomes an active system.[35]

Some more general types of self-phasing arrays have been proposed by a number of workers. A number of physical embodiments

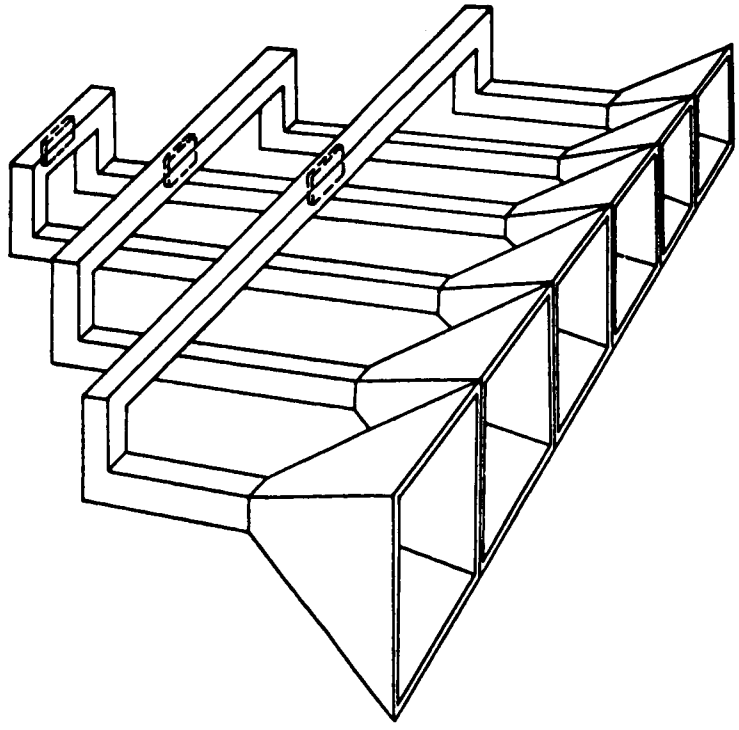


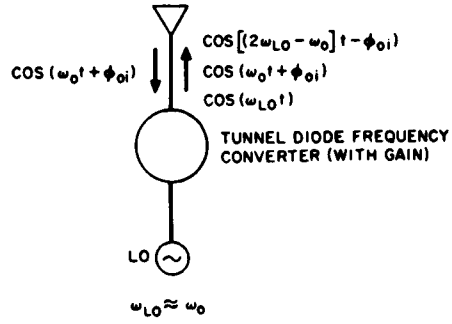
Figure 4-15. Van Atta array.[34]

may be realized; the techniques of Rutz-Philip,[36] Sichelstiel et al.,[37] Cutler et al.,[38] and Pon[39] are sketched in Figure 4-16. The principle of operation of all the systems shown in the figure is essentially the same. The phase of the signal incident at each element is reversed by the electronic circuitry. This is just the condition required to steer a beam in the direction of the incident radiation. The signals at various points in the system are shown on the diagrams of Figure 4-16.

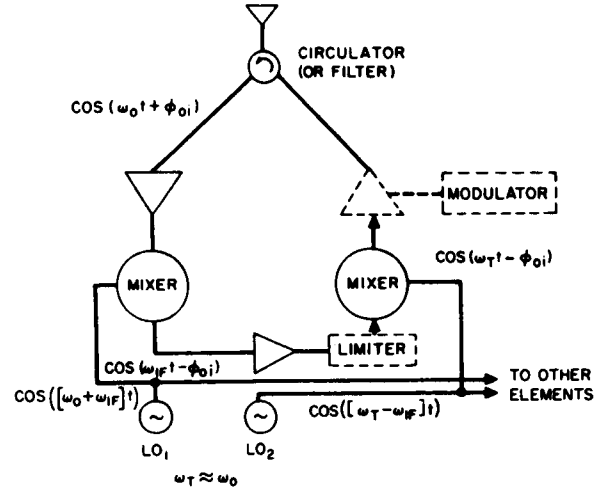
In addition to the arrays discussed above which send energy back in the direction of an incident pilot signal, there are systems which automatically receive information from the direction of an incident signal with full array gain. They may also radiate different information back in the same direction, or in more advanced systems, in some other direction dictated by a pilot signal. Electronic techniques similar to those already mentioned or adaptive techniques may be used. Adaptive techniques utilize phase-locked loops to accomplish the required antenna phasing rather than phase inversion by mixing. Figures 4-17a and 4-17b illustrate techniques, using only mixing, that form a receiving beam as well as a retransmitting beam. Figure 4-17a shows a retrodirective system and Figure 4-17b a redirective system.

Phase-locked loops have been finding increasing application in recent years as phase synchronizing devices for an arbitrary distribution of radiating elements. An array of elements employing this type of phase control is called an electronically adaptive antenna system. A receiving antenna system of this type, unlike passive arrays, contains active elements that automatically adjust the electrical phases of the signals received by the array elements to obtain antenna directivity. These arrays can be made retrodirective. A configuration of this type is illustrated in Figure 4-18.

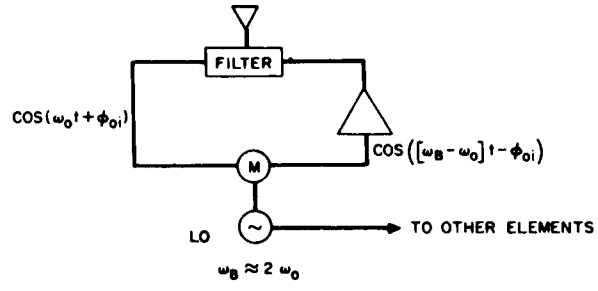
The characteristics of phase-locked loops are well documented and, though their construction is more complex than that of the mixers and filters required for the phase inversion systems, the performance of systems using them may be greatly superior to the performance of the simpler phase-inversion systems. This comparison is especially true when the individual modules are operating at low signal-to-noise



(a) SYSTEM OF RUTZ-PHILIPP (1964)



(b) SYSTEM OF SICHELSTIEL ET AL. (1964)



(c) SYSTEM OF CUTLER ET AL. (1963) AND C. Y. PON (1964)

Figure 4-16. Retrodirective array techniques.

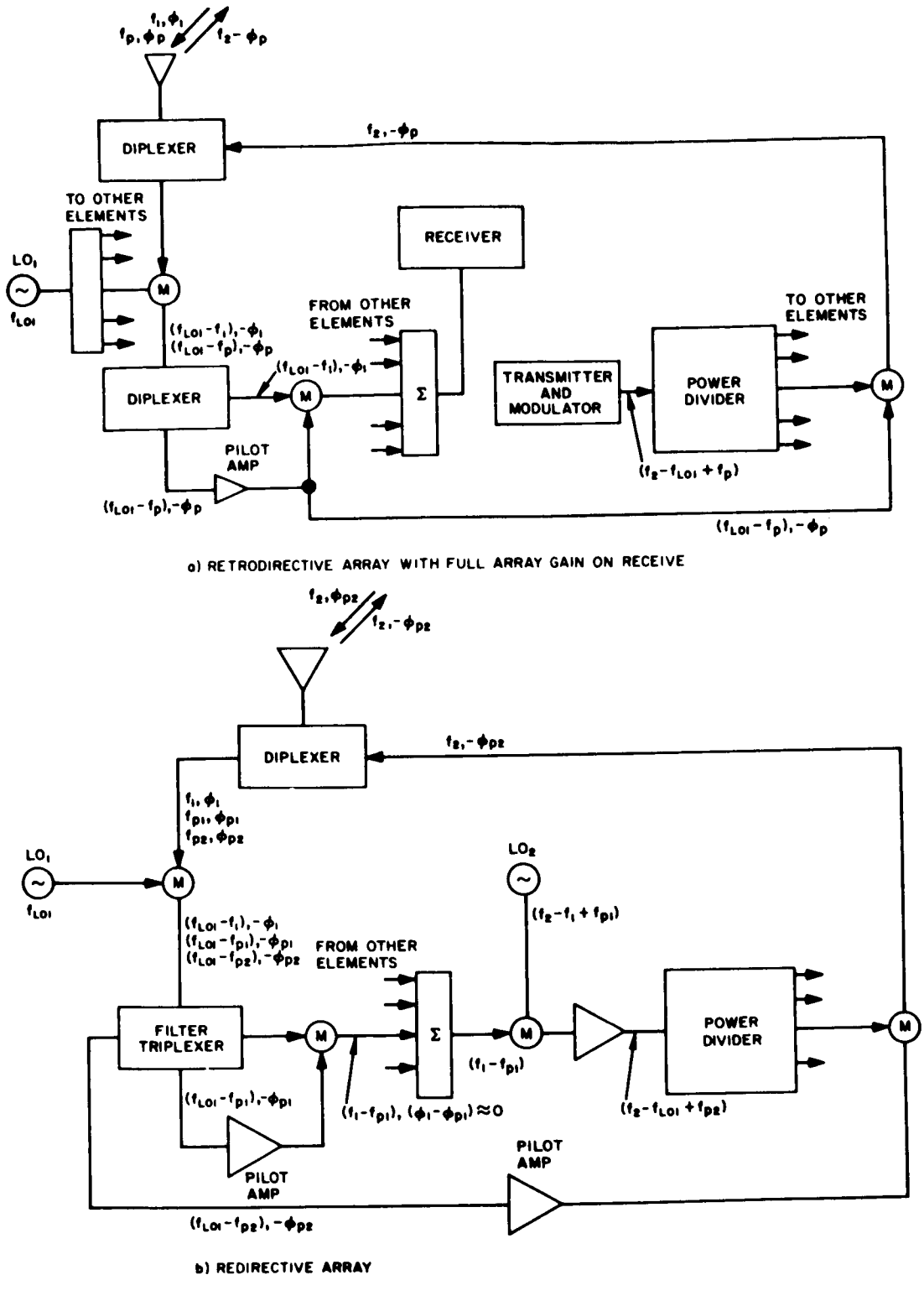


Figure 4-17. Electronic beam-forming techniques.

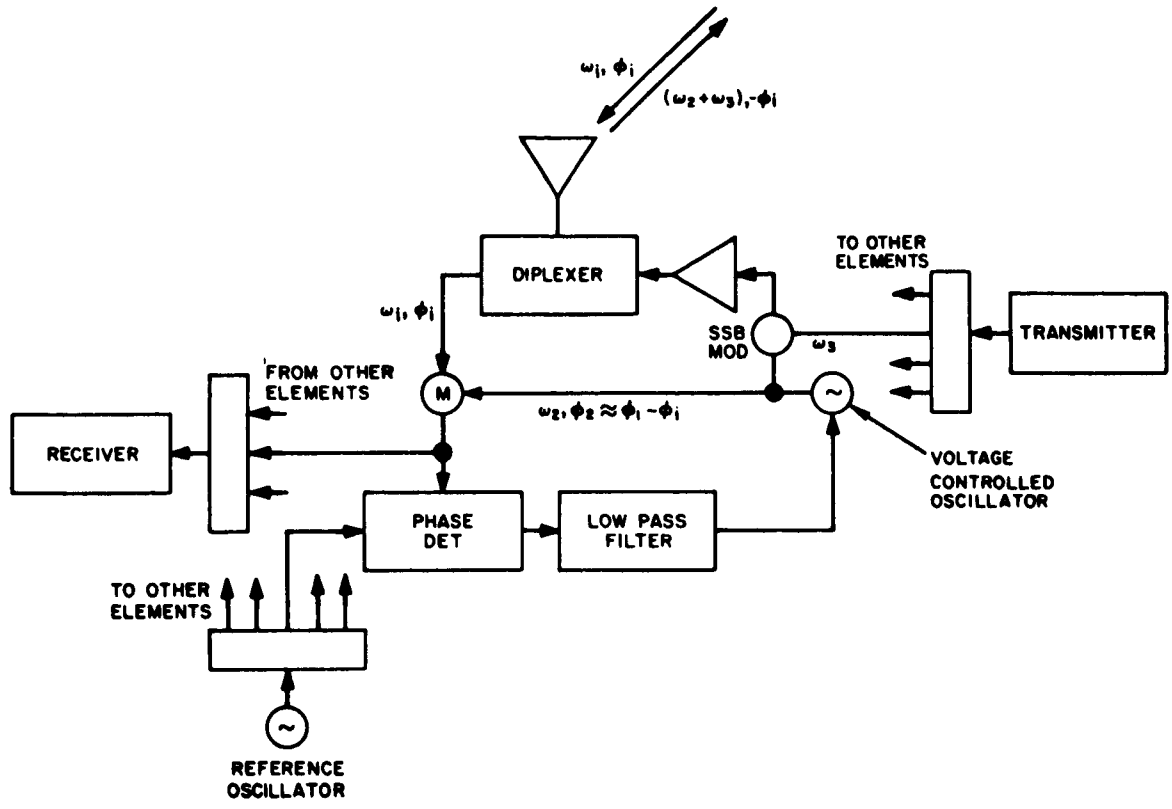


Figure 4-18. Adaptive retrodirective antenna module.

ratios. Because of this prospect of better performance and from a purely technical point of view, the system using phase-locked loops appears the more desirable for reception. However, from an economic point of view it might be more advantageous to put additional power into the signal transmitted from earth and to use the simpler phase-inversion scheme than to use the more complicated phase-locked receiving system on the spacecraft. A careful comparison of relative performance versus cost is required to determine to what extent such a tradeoff might be desirable, or if it is even reasonable, for the distances involved in interplanetary missions.

The same techniques that give beam steering on reception also provide the beam steering for the retransmission process. The system

that uses phase inversion by mixing translates the received pilot signal in frequency and retransmits it. Since there is receiver noise superimposed on the received pilot signal, the bandwidth of the filters used in the retransmission process will determine the relative magnitudes of the retransmitted pilot signal and this receiver noise. Unless adaptive filters which can have very narrow noise bandwidths are used, the noise relative to the retransmitted pilot signal can be appreciable when the incident pilot signal level is small. For the system that uses phase-locked loops, there is essentially no additive noise on the retransmitted signal, but there is phase noise that is introduced by the jitter in the voltage-controlled oscillators of the phase-locked loops. When the loops are operating properly, however, the phase noise can be expected to be small. Experience with probes such as Mariner IV indicate that the phase noise on a single channel is acceptable, at least for the bit rates used (about 8-1/3 bits per second at Earth-Mars distances).

A much greater bit rate (10^6 per second) is considered in the present study. It should be pointed out, however, that the high bit rate will probably not be required on the earth-to-spacecraft link (up-link) immediately, since probably only commands would be transmitted on this link. The spacecraft-to-earth link would require the high-bit rate. Therefore, in the immediate future, the up-link bit rate may not be greatly different than those of previously employed systems. Manned missions would require considerably increased up-link data rates; and if a communication up-link using television bandwidths were required, then 10^6 bits per second on the up-link would be a minimal requirement.

The type of coding and modulation-demodulation system employed in the communications has an effect on the reliability of received messages. A study of the effects on error probabilities of using self-phasing and adaptive arrays is presently underway on Contract NAS 2-3297.* In this study probabilities of errors are being determined

*Study of Applications of Retrodirective and Self-Adaptive Electromagnetic Phase Controls to a Mars Probe.

for adaptive arrays that use PSK, DPSK, FSK (coherent and incoherent), and ASK (coherent and incoherent) modulation. The analysis includes the effects of additive receiver noise and of phase errors introduced by imperfect phase tracking of the phase-locked loops. The results of these studies, when completed, should be applied to the mission of interest in this study. The studies do not include the effects of fading and multipath on the probability of error. These effects should be investigated since in the near neighborhood of planets, the interference between the direct signal and reflections from planetary surfaces can have adverse effects on the correct decision in determining which of the incoming signals is being received at any instant of time. A method must be found to sort out these signals so that the equivalent of the signal coming via the direct path is processed.

Experimental Systems. A number of experimental models have been built that demonstrate self-steering techniques. Several significant examples are given in the following paragraphs. The listing of examples, however, is not exhaustive.[40]

Van Atta Arrays. An experimental model of the Van Atta type using two arrays has been built and tested,[41] but the arrays were not scaled in size to offset the effect of frequency shift. The model (Figure 4-19) uses two interleaved arrays of orthogonally oriented printed dipoles and employs tunnel-diode amplifiers and mixers. The orthogonal orientation of the elements in the two arrays provides isolation between received and retransmitted signals. In this system, the information signal, which has been obtained from some other source, is impressed on the reradiated signal by introducing it on the local oscillator signal at frequency Δf. The experimental array has nine active elements and operates at a nominal receiving frequency of 2.0 GHz and transmitting frequency of 2.15 GHz. The r-f bandwidth is approximately 120 MHz, and the measured array gain is reported to be 14 db. The array was operated under conditions of both single and multiple access.

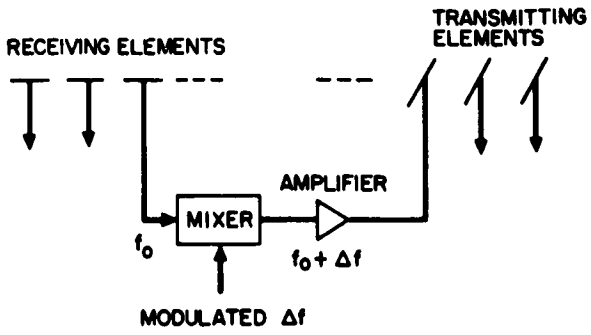


Figure 4-19. Active Van Atta system with separate arrays for reception and retransmission.

A second experimental version of the Van Atta array has been described briefly by Gruenberg and Johnson.[42] It consists of a 25-element Van Atta array in which 24 conjugate elements are pulse-code modulated by on-off switches. The switching modulates the reradiated signal. The modulation comes from the decoded signal received by the 25th (center) element. This receiving element accepts the modulated signal from a distant transmitter. A separate receiving station then beams an unmodulated signal to the array. This signal is modulated by the array switches and directed back to the receiving station. A diagram of the system is illustrated in Figure 4-20. No experimental results were given in the reference.

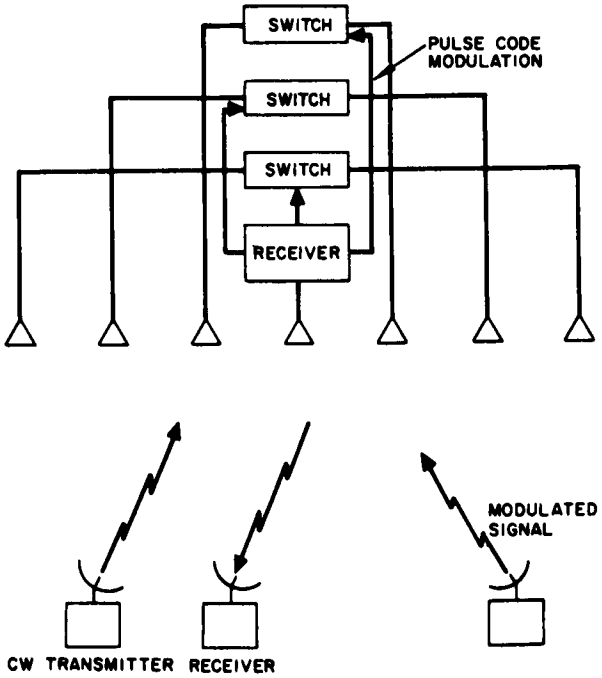


Figure 4-20. Modulated Van Atta array.

Multiple Beam Antennas. A multiple-beam antenna system that implemented a Transdirective self-steering concept was designed and built in a breadboard model by the Hughes Aircraft Company Antenna Department on Contract NAS 5-3545. The system had two 4-by-4 arrays of helical elements, and two separate RF sections, one for transmitting and one for receiving. Similarly, there were two matrices for the formation of the multiple beams. A detailed discussion of the operation of this type of self-steering system and a performance evaluation is available in Reference 33.

Phase Inversion by Mixing. Sichelstiel et al.[37] have described a retrodirective linear array operating at X-band (Figure 4-16b). The array consists of 25 horn radiators arranged horizontally. Much of the RF circuitry is waveguide. The 10.7 MHz IF contains a crystal filter to eliminate unwanted frequencies. The incident RF frequency was 9.345 GHz and the retransmitted frequency was 9.385 GHz. A number of radiation patterns were shown for 50-wavelength and 100-wavelength arrays. The patterns were taken with the source of incident radiation at different ranges and at on-axis and off-axis positions. The focusing properties were illustrated when a dielectric sheet was partially interposed between the source and the array. The operation for extended sources was also investigated.

A second experimental retrodirective array is described by C. Y. Pon, Figure 4-16c.[39] This array consists of 4 slot-fed dipoles, each with a duplexer and a balanced mixer. The mixers are in waveguide and coaxial line while the duplexer is in stripline. The incident frequency was 2.95 GHz, the local oscillator frequency was 5.75 GHz, and the reradiated frequency was 2.80 GHz. Experimental patterns are presented in the paper for several angles of incidence of the signal. The element spacings were such that grating lobes were present in all patterns.

One applicable redirective beam-steering technique forms high-gain beams from an array of elements for a transmitting mode and uses high-gain beams from another antenna for a receiving mode. An

engineering model of a self-steering microwave transponder implementing this concept is being designed, constructed, and evaluated at the Hughes Aircraft Company. A discussion of this system is given in Appendix D.

Experimental Phase-Locked Loop Array. An eight-channel antenna array utilizing phase-locked loop principles has been simulated.[43] A block diagram of the simulator is shown in Figure 4-21. The signal picked up by the individual antennas was simulated by a 1-Khz sinusoid, and the zero-voltage voltage-controlled oscillator (VCO) frequencies were approximately 1300 Hz, giving an IF frequency of 300 Hz. The phase shifts in each antenna were simulated by resolver phase shifters. To measure the system adaptiveness, the acquisition time as a function of signal level was measured (Figure 4-22). The noise bandwidth of the loop was 3.2 Hz, the damping constant was equal to 2, and the so called open-loop gain was $K = 2260\ \text{sec}^{-1}$. The sum signal was held constant as the input signal was changed.

It was found that the acquisition time decreased as the signal strength increased, thus indicating an increase in the noise bandwidth. Oscillograms of the VCO phase jitter (Figure 4-23) also indicated an increase in the noise bandwidth with signal strength by the presence of higher frequency components in the oscillograms for large signal strength.

Conclusions. Of the general types of antennas discussed, only the reflector and feed antenna has been treated as a presently available device. The theoretical and experimental performance characteristics of this type of antenna are available, and it has been used extensively for various applications including long range communications. The phased array and the self-steering array, on the other hand, represent relatively new techniques, and their applicability to space communication requires more study. Several versions of these two antennas have been studied experimentally in the laboratory by different groups of investigators, but the efficiency and reliability of the antennas in a space environment or as part of a large ground system has not been

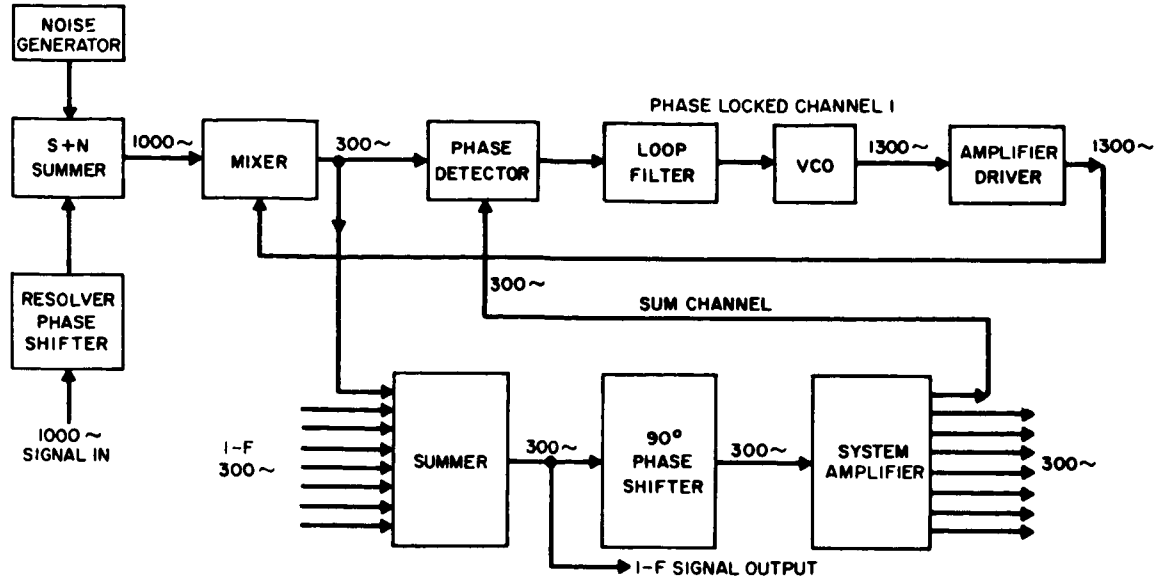


Figure 4-21. Block diagram of the active adaptive antenna array system simulator. [43]

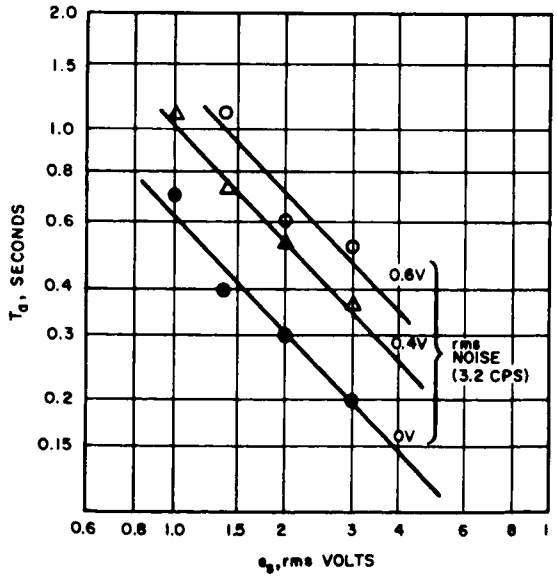


Figure 4-22. Acquisition time of the 8-channel antenna array simulator vs. mixer output voltage. [43]

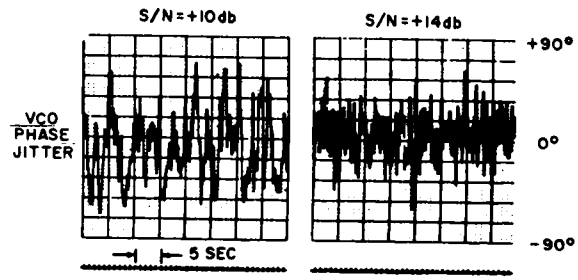


Figure 4-23. Oscillograms showing the adaptiveness of the phase-locked system. [43]

thoroughly evaluated. The increased flexibility of operation offered, however, indicates that phased-array techniques merit further development, particularly in relation to deep space communication applications.

4.5.5 Ground-Based Antennas

General. From the discussion of system performance in Section 3.2, it is apparent that one of the promising frequencies of operation is in the microwave region and probably in the S-band. In subsection 4.5.1, the discussion points out that at 2.3 GHz dimensions on the order of 600 to 1000 feet or more are probably realistic sizes to consider for the requirements imposed by the high data rate needed for a deep space communication system. Various possible antenna types are discussed in 4.5.4. Consideration of the practical difficulties that have been encountered in the development of reflector antennas of this size - as at Sugarloaf, West Virginia, and at Arecibo, Puerto Rico - tend to limit the choice to some type of distributed aperture and possibly to a phased array. This selection merits further consideration, based on investigations of the available and potential technology of the necessary components, as a function of the aperture size and type of elements. The array size and the requisite number and kinds of components must be optimized in relation to the losses associated with the transmission lines and the effective antenna noise temperature.

Scanning arrays consisting of a large number of elements, usually paraboloids, dipoles, or spirals, have been used for radio astronomy. At the 1965 AIAA meeting in San Francisco, California, it was proposed that 10,000 reflector antennas, each 100 feet in diameter, be used to cover an aperture of six miles in diameter for a radio telescope. The JPL study[10] points out that an array (about 10 elements) of 210-foot antennas could be considered economically feasible for very

deep space communication systems with the advent of Nova boosters or multiple-rendezvous Saturns and spacecraft transmitter power in the range of 100 w to 1 Kw.

In one possible implementation of such a large array, it could be subdivided into subarrays of, say, 100 feet by 100 feet. These subarrays would then be elements of the array that could be scanned electronically or mechanically or by a combination of the two methods to optimize the gain as a function of scan angle. The subarrays could be made up of individually discrete radiators or of continuous apertures as has been suggested for the radio telescope. Further investigations must be made to determine the optimum combination of the element factors and the proper array factor that will prevent grating lobes from appearing within the required angular scan range. In addition, a detailed study of the components associated with feeding each subarray and the basic radiators must follow to arrive at a set of feasible tradeoffs that may be made between the element and the array characteristics for various possible operating frequencies.

Self-steering techniques as described in Section 4.5.4 appear to hold some promise for the highly directive ground-based antenna that must respond to the relatively weak signals from a distant spacecraft. These techniques, in principle, can alleviate the pointing difficulties of high-gain antennas. It should be observed that only a very narrow information bandwidth is required for pointing accuracy. This low data rate requirement for beam pointing is an especially significant factor in determination of the size of the subarrays in a large ground-based system. Further specific study is needed to determine the amount of self-steering that can be accomplished and the associated cost in component complexity. In addition, such techniques are desirable for practical construction of electrically large arrays at the microwave region. At these frequencies, manufacturing tolerances and atmospheric scintillation effects may cause a deterioration of the antenna's characteristics due to the decorrelation of the signals at the various

elements. Again, in principle, a self-steering array should automatically correct for these effects and allow the maximum amount of the full gain of the array aperture to be realized. The details of these techniques and their developmental progress to date have been discussed in 4. 5. 4. Questions about their reliability and ultimate accuracy still remain to be answered by more intensive research and development.

Implementation of the possible antenna techniques and the optimum choice of antenna type are both contingent upon the practical solution of several problems of varying degrees of technical complexity. In the discussion of Section 4. 5. 3 and Appendix A, on the subject of noise due to the presence of the sun in the sidelobes, the duration of the period during which communication may be lost is related to the level of the sidelobes of the antenna pattern. The computations presented were made on the basis of a continuous aperture illumination. In any array of mechanically positioned subapertures, the aperture distribution will be discontinuous, with the appearance of grating lobes as a consequence. This problem is treated briefly in a following subsection. It must be recognized that, although the emphasis therein is with respect to mechanically pointed subsections of the aperture, the problem of grating lobes is not a trivial one with electronically scanned arrays. Such lobes occur whenever there is any significant discontinuity in the phase or amplitude of the aperture illumination.

Another question of interest is the determination of the appropriate size of the subaperture of a large array. In addition to the matter of cost touched upon in Section 4. 5. 3, other factors can be significant in bounding the subaperture size. A strong noise source may be present in the pattern of the subaperture alone at greater angles off the axis of the main beam of the whole array than for the main beam itself. The effect on the overall antenna temperature of the presence

of the sun in the subaperture pattern has been analyzed in part for an adaptive array and the analysis is presented below. A numerical example is given and the results tend to show that the presence of sun in the subaperture pattern does not preclude the use of this kind of antenna.

When an array becomes sufficiently large, differences in propagation time of the incoming signal between one side of the array and the other can become significant relative to the time involved in changes in the modulation that conveys the desired information. Time delay compensation becomes necessary, although the type of compensation depends on the details of array implementation. The relative time delays that are permissible across portions of the overall array may set an upper limit on the subaperture size. Consideration is given to time-delay compensation in a third subsection below. A discussion as to the appropriate subaperture size is then given with a tentative choice indicated in a fourth subsection. A discussion of distribution systems and phase shifters that make up types of subapertures and an indication as to the most promising kind completes the examination of ground-based antennas.

Grating Lobes in Large Arrays. The formation of grating lobes in large array of parabolic reflectors constitutes a serious difficulty for which no generally satisfactory solution has yet been developed. The problem can be visualized if the array pattern is considered as the product of an element pattern and an array factor. The element pattern consists of the radiation pattern produced by a parabolic reflector, while the array factor is the pattern of an array of isotropic radiators and is a two-dimensional grating lobe pattern. It is steered electronically while the element pattern is directed by the mechanical movement of the individual dishes. In the ideal case the element pattern and a single lobe of the array factor will both point in the desired direction. Multiple beams appear, however, when more than one grating lobe falls

within the main beam of the element factor; this condition occurs when the array spacing is substantially greater than the diameter of the sub-apertures.

In illustration, a square array of 14 elements on a side is considered. Each element is a parabolic reflector antenna 100 feet in diameter with a 26-db Taylor amplitude distribution. The spacing between reflector centers is 150 feet; this distance permits scanning to 48 degrees from broadside without interference between adjacent array elements. The layout of elements is illustrated in Figure 4-24. The radiation pattern of the array is given by

$$P(\theta\phi) = \left[P_a(\theta\phi)\right] \left[P_\epsilon(\theta\phi)\right] \tag{12}$$

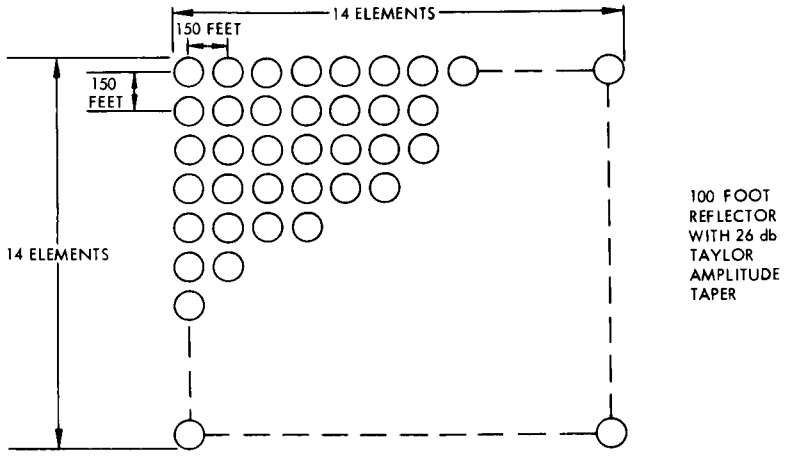


Figure 4-24. Layout of array elements.

where $P_a(\theta\phi)$ is the array factor and $P_\epsilon(\theta\phi)$ is the element factor. The element factor is, of course, the pattern produced by the reflector, and because the reflector is mechanically steered, it has the same shape for any pointing angle. The array factor is of the form,

$$P_a(\theta\phi) = \left[\frac{\sin\frac{N\pi a}{\lambda}(\sin\theta\cos\phi-\alpha)}{\sin\frac{\pi a}{\lambda}(\sin\theta\cos\theta)}\right]^2 \left[\frac{\sin\frac{N\pi a}{\lambda}(\sin\theta\sin\phi-\beta)}{\sin\frac{\pi a}{\lambda}(\sin\theta\sin\theta-\beta)}\right]^2 \qquad (13)$$

where N = the number elements on a side

λ = wavelength

α = phase delay between elements in the $\phi = 0$ plane

β = phase delay between elements in the $\theta = 0$ plane

a = element spacing

The array factor is steered by changes in α and β which are the progressive phase shifts between elements in orthogonal planes.

It can be seen from this equation that the beamwidth of the array factor, which is essentially identical to the beamwidth of the array, increases as the beam is scanned from the broadside position. However, the gain of the system does not decrease as might be expected because the relative amplitude of the grating lobes decreases as the beamwidth increased. This tradeoff only occurs for pointing angles small enough so that one reflector does not block another.

The element and array factors for the broadside beam position are sketched in Figure 4-25 and the overall array pattern is shown in Figure 4-26. The patterns presented show only the two principal planes; a three-dimensional view is given in Figure 4-27. It may be seen that at least eight significant grating lobes are present in the pattern produced by the array of Figure 4-24. These lobes have an amplitude on the order of 4 db below the main beam and, in the two principal planes, are located approximately 0. 163 degree away from the main beam. If the array is scanned away from broadside in one of

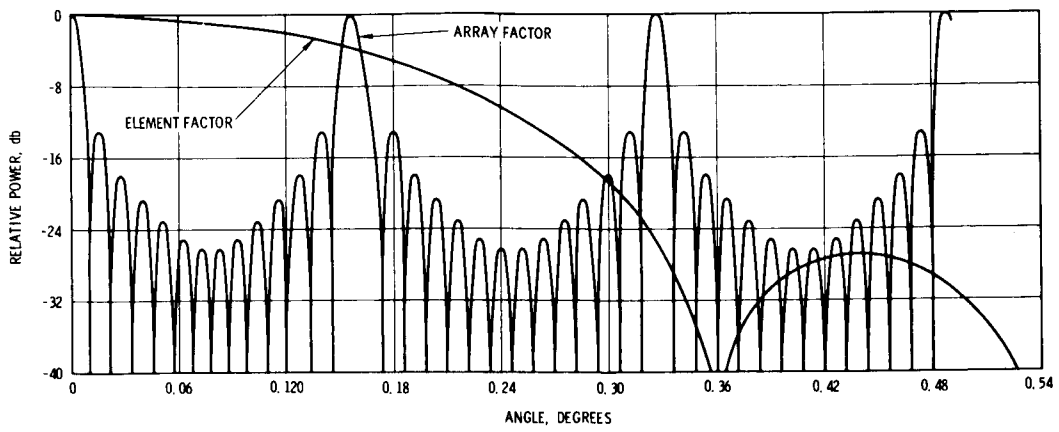


Figure 4-25. Element and array factors for the array of Figure 4-24.

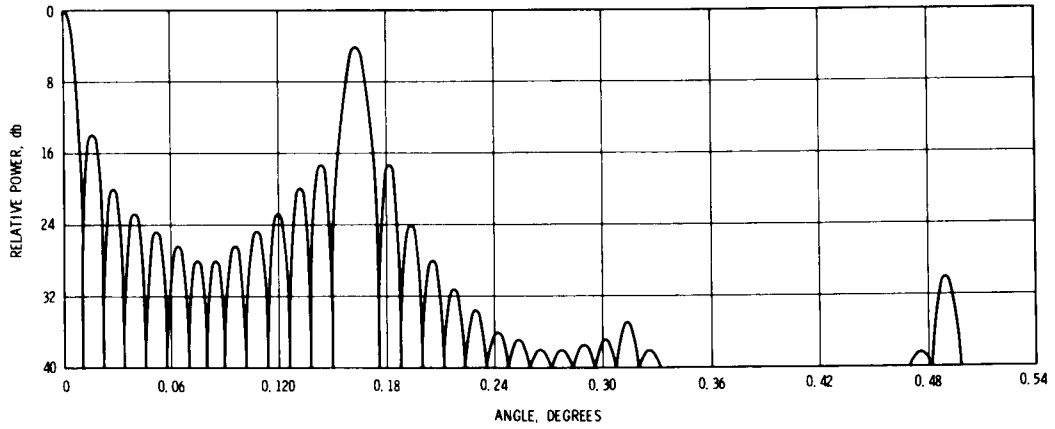


Figure 4-26. Radiation pattern for the array of Figure 4-24.

the principal planes, two effects will be observed in that plane. First, the angular separation of the grating lobes from the main beam will increase inversely with the cosine of the scan angle, and second, the amplitude of the grating lobes will decrease. These effects are noted because as the array factor is scanned, the lobe separation increases as a result of the decrease in the effective or predicted separation of the elements. The element factor beamwidth, however, is invariant with scan angle and, consequently, the grating lobes move farther out along the element pattern with a corresponding decrease in amplitude compared with the principal or main beam.

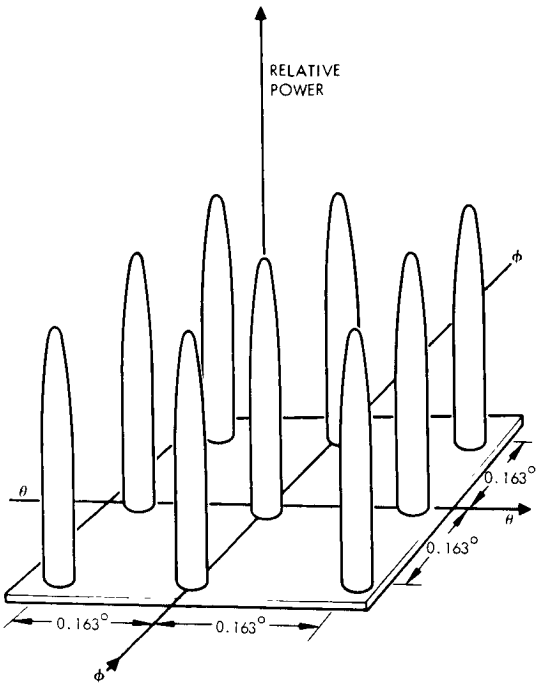


Figure 4-27. Three-dimensional plot of principal grating lobes surrounding the main beam.

Each grating lobe is accompanied by a sidelobe structure which is determined by the amplitude weighting applied to the array. For the example given, in which a uniform distribution is assumed, the sidelobes follow a $\frac{\sin X}{X}$ distribution with the first sidelobe at the usual 13 db level. In practice, the sidelobe structure around each grating lobe can be reduced by application of any of the standard amplitude tapers used in the achievement of low sidelobe levels.

The grating lobes are only objectionable when communication is desired with a vehicle that is nearly eclipsed by a noise source, such as the sun. In this case one of the grating lobes on the beam side nearest the noise source can probably be pointed at the vehicle. The effective gain will, of course, be somewhat reduced, but at least, a situation in which one of the stronger grating lobes points directly at the noise source will be avoided.

The spacing of the grating lobes from the main beam can be increased by a decrease in the separation of the parabolic reflector antenna elements. However, if this spacing is decreased, the diameter of the reflectors must also be decreased so that the effective scan range can be maintained, while at the same time more array elements must be added to meet the gain requirement. The end result will be a broader element pattern which in turn will ensure that the grating lobes will have essentially the same amplitude relative to the main beam. The beamwidth of both the main beam and the grating lobes will, for all practical purposes, remain the same as long as the overall array dimensions remain unaltered. The fine grain structure around the various lobes will change, however, as more elements are added.

If the spacing between the elements is increased, and the diameter of the reflectors is increased correspondingly, the grating lobes will be moved in closer to the principal beam. Once again the relative amplitude and beamwidth of all the grating lobes should remain essentially constant.

A possibility exists that the grating lobes adjacent to the principal beam may be reduced in amplitude by the use of random spacing among the array elements. However, it is anticipated that the selection of such a design will prove to be an extremely difficult problem. Intuition suggests that, at best, a few large grating lobes may be exchanged for a relatively large number of moderately sized lobes. In addition, it is quite likely that the obtaining of any real improvement over more than a limited range of scan angles would not be feasible. One problem which arises, for example, in a phased array of unevenly spaced elements, is that the phase distributions in the two principal planes are no longer separable or independent functions. Therefore scanning in one plane may affect the amplitude and position of the grating lobes in the other plane.

Another means of suppressing the grating lobes might involve the use of an auxiliary array that could be steered and phased to cancel out any given lobe. A major difficulty that might be anticipated from such a scheme would be the obtaining of sufficient gain from the auxiliary array.

Temperature of Adaptive Arrays Due to the Sun. The problem of concern is whether the effect of the noise generated by the sun, and appearing at the output of a subaperture of a large adaptive array, significantly degrades the operation of the adaptive array when its beam is pointing near the sun.

Since the large aperture is composed of subapertures that are phased together by means of phase-locked loops located at each subarray, the sun will affect the subaperture temperature differently than the temperature of the total antenna. As the subapertures are scanned near the sun, their temperatures will rise and the signal-to-noise ratio in the tracking loops will decrease. As this ratio decreases, the phase errors of the individual subaperture signals will increase.

Since the noise due to the sun at the different subapertures comes from a common source, there will be some correlation among these noises. If the noise was correlated at the elements and if the phase-locked loops were all identical, the phase errors introduced by the noise in each phase-locked loop would be a predictable function of the position of the associated element in the array and of the angular positions of the signal and noise sources. It appears that under most circumstances the aggregate effect of the errors in the loops on the errors in the communication channel is much like that of the case in which the noise is uncorrelated. Special conditions need to be investigated to determine worst case performance.

In the case in which there is negligible correlation among the noises in the individual tracking loops, the overall array pattern would be degraded somewhat, and as the beam approaches the sun, the effective noise temperature of the total array would increase more rapidly than it would if the phasing were insensitive to the subaperture noise temperature. To determine how the subaperture noise temperature affects the total array temperature, an analysis must be made of the phase errors in the tracking loop as a function of the signal-to-noise ratio in that loop. Then the phase error must be related to the noise temperature of the array. Such an analysis allows a determination of the least distance to the sun at which a spacecraft can be viewed.

An analysis of the type considered, in which the correlation of the noises is accounted for, has not yet been carried out. However, an analysis was made that assumes statistical independence of the noise, and hence, of the phase errors, from subaperture to subaperture. This assumption should lead to somewhat optimistic results as compared with the case in which the noise is perfectly correlated.

Analysis. The satisfactory operation of the communication channel will require the maintenance of a minimum operating value of energy per bit, E, to noise spectral density, $2N_o$, for an acceptable probability of error. If the value of $E/(2N_o)$ falls below this value by more than a prescribed margin, the communication channel will no longer operate satisfactorily.

When the antenna beam is not directed near the sun, the contribution of the sun to the noise temperature of the antenna may be neglected in comparison with the temperature of the rest of the system, T_r. Under these conditions, the system will operate with a prescribed value of signal-to-noise ratio, $(E/2N_o)_o$, where

$$E = P_o T_b \tag{14a}$$

$$2N_o = k T_r \tag{14b}$$

where

P_o = the average power received in the communication signal

T_b = the bit duration

k = Boltzmann's constant

T_r = system temperature

Now, α is the ratio of P_T, the power in the tracking signal, to P_o, and B_ℓ is the tracking loop bandwidth. Then, the normal operating signal-to-noise ratio in the tracking loops at the individual elements may be written

$$\left(\frac{P_T}{P_{N\ell}}\right)_o = \frac{\alpha}{\eta M B_\ell T_b} \left(\frac{E}{2N_o}\right)_o \tag{15}$$

where η is the array aperture efficiency and M is the number of subapertures.

This signal-to-noise ratio in the tracking loop controls the phase error of the individual subaperture. It has been shown [44] that for loop signal-to-noise ratios greater than about 9 db, the mean squared phase error in the loop is just the reciprocal of the loop signal-to-noise ratio when the loop is a proportional-plus-integral control loop. It has also been shown [44] that the probability density function of the phase errors is approximately Gaussian for high signal-to-noise ratios.

Thus

$$P(\psi) \approx \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\psi^2}{2\sigma^2}\right) \tag{16a}$$

where

$$\sigma^2 \approx \left(\frac{P_{N\ell}}{P_T}\right) \tag{16b}$$

When the array is steered toward the sun, the noise temperature of the subapertures will rise because of the contribution from the sun, T_{es}^{e}. The system temperature at the subaperture then becomes $T_r + T_{es}^{e}$, and the loop signal-to-noise ratio may be written as

$$\left(\frac{P_T}{P_{N\ell}}\right) = \left(\frac{P_T}{P_{N\ell}}\right)_o \frac{T_r}{\left(T_r + T_{es}^{e}\right)} \tag{17}$$

and as a result of the reduced signal-to-noise ration the rms phase errors at the subapertures increase. Consequently, the total aperture gain will be reduced, and the pattern sidelobe level will increase. The increased sidelobe level will result in an increase in the total array

effective noise temperature due to the sun, T_{es}^{a}. These quantities can be evaluated in a straightforward manner for the mean power pattern when independent gaussian amplitude and phase errors are assumed.

Gain of Mean Power Pattern. An expression for the gain of the mean power pattern is derived first. The excitation coefficients of a two-dimensional rectangular array of M identical elements are denoted as A_{mn}. In general there will be amplitude errors Δ_{mn} and phase errors ψ_{mn}. The field pattern of the array may be written as

$$E = f(u-u_o,\ v-v_o) \sum_m \sum_n A_{mn}(1 + \Delta_{mn}) \exp\left(jk\left[m(u-u_o) + n(v-v_o)\right]\right)$$

$$\exp(j\psi_{mn}) \tag{18}$$

where

f = the element pattern normalized to unity at its peak

u = sin θ cos φ

v = sin θ sin φ

The power pattern is then proportional to EE*. If the amplitude and phase errors are assumed to be independent gaussian random variables with variances $\overline{\Delta^2}$ and σ^2, respectively, then the mean power pattern may be computed in a straightforward manner to be

$$\overline{P(u-u_o,\ v-v_o)} = P_o(u-u_o,\ v-v_o)\exp(-\sigma^2) + \left| f(u-u_o,\ v-v_o) \right|^2$$

$$\cdot \left[1 - \exp(-\sigma^2) + \overline{\Delta^2}\right] \sum_{mn} \left| A_{mn} \right|^2 \tag{19}$$

The gain of this mean pattern is given by

$$G = \frac{4\pi P}{\iint \frac{P\ du\ dv}{\sqrt{1-u^2-v^2}}} \tag{20}$$

For narrow beam antennas near broadside, Equation (20) becomes

$$G = \frac{4\pi P}{\iint P \, du \, dv} \tag{21}$$

When Equation (19) is substituted into Equation (21) and the integration is carried out, the following expression for gain results.

$$G \approx \eta \frac{4\pi A}{\lambda^2} \exp(-\sigma^2) \frac{\left\{1 + \frac{1}{\eta M}\left[\exp(\sigma^2)(1 + \overline{\Delta^2}) - 1\right]\right\}}{\left[1 + \overline{\Delta^2}\right]} \tag{22}$$

In the derivation of Equation (22), it has been assumed that the array subapertures are uniformly illuminated and use has been made of the relation

$$\frac{P_o(0, 0)}{\sum |A_{mn}|^2} \equiv \eta M$$

where η is the aperture efficiency of the array excitation and M is the number of elements.

Temperature of Array in Terms of Element Temperatures. The effective noise temperature of an antenna is given by

$$T_e = \frac{1}{4\pi} \iint T \, G(u, v) \frac{du \, dv}{\sqrt{1 - u^2 - v^2}} \tag{23}$$

where T is the temperature of the observed sources. When the expression of Equation (22) is employed in Equation (23) and the various terms identified, the effective array temperature is shown to be

$$T_{es}^{a} = \frac{T_{eso}^{a} + \left[1 + \overline{\Delta^2} - \exp(-\sigma^2)\right] T_{es}^{e}}{1 + \overline{\Delta^2}} \tag{24}$$

where the effective array temperature in the absence of excitation errors, T_{eso}^{a}, is

$$T_{eso}^{a} \equiv \frac{1}{4\pi} \iint G_o \frac{T \, du \, dv}{\sqrt{1 - u^2 - v^2}} \tag{25}$$

$$= \frac{\eta A}{\lambda^2} \iint \frac{T \, P_o(u - u_o, v - v_o) \, du \, dv}{P_o(0, 0) \sqrt{1 - u^2 - v^2}}$$

and the effective subaperture temperature, T_{es}^{e}, is

$$T_{es}^{e} = \frac{A}{M \lambda^2} \iint T \frac{\left| f(u - u_o, v - v_o) \right|^2 du \, dv}{\sqrt{1 - u^2 - v^2}} \tag{26}$$

It can be seen that in the absence of errors, T_{es}^{a} equals T_{eso}^{a} but that when mean square errors are appreciable, T_{es}^{a} approaches T_{es}^{e}.

Numerical Example. As an illustration of the analysis, the following system is considered. The antenna is assumed to have a square aperture 2000 feet on a side, composed of 400 contiguous, square subapertures 100 feet on a side. Frequency of operation is 2.3 GHz. The required operating communication signal-to-noise ratio, $(E/2N_o)_o$, is assumed to be 10 db. The bit rate is 10^6 bits per second, i.e., $T_b = 10^{-6}$ second, and the loop noise bandwidth, B_ℓ, is assumed to be 10 cycles per second. The value of α is taken to be 0.10. It can be found that

$$\sigma^2 \approx 0.263 \times 10^{-2} \left(\frac{T_r + T_{es}^{e}}{T_r} \right) \tag{27}$$

To complete the analysis, the values of T_{es}^{e} and T_{eso}^{a} are required as a function of pointing direction relative to the sun. The pattern of uniformly illuminated square aperture of side $\sqrt{\frac{A}{M}}$ is given by

$$G = \frac{4\pi}{\lambda^2} \frac{A}{M} \frac{\left[\sin \frac{\pi}{\lambda}\sqrt{\frac{A}{M}}\ (u - u_o)\right]^2}{\left[\frac{\pi}{\lambda}\sqrt{\frac{A}{M}}\ (u - u_o)\right]^2} \frac{\left[\sin \frac{\pi}{\lambda}\sqrt{\frac{A}{M}}\ (v - v_o)\right]^2}{\left[\frac{\pi}{\lambda}\sqrt{\frac{A}{M}}\ (v - v_o)\right]^2} \qquad (28)$$

For the antenna considered, $\sqrt{\frac{A}{M}} = 100$ feet and $\left(\sqrt{\frac{A}{M}}\right)\frac{1}{\lambda} = 234$. The temperature due to the sun was evaluated on a GE-265 time sharing computer as a function of angular separation between the element pattern peak and the edge of the sun. The frequency considered was 2.3 GHz. The resulting curves of temperature versus angular separation are shown in Figure 4-28.

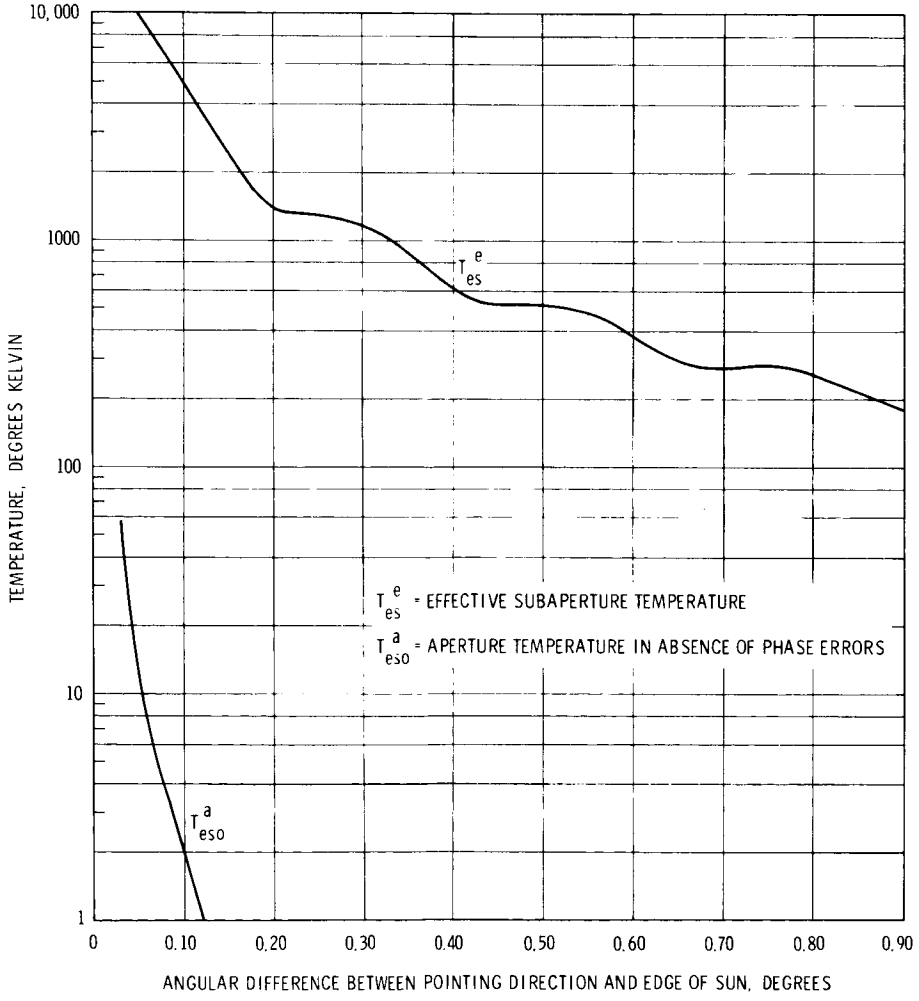


Figure 4-28. Effective temperatures without errors.

For the calculation of the array temperature, it was assumed that the subapertures were uniformly illuminated in amplitude and were phased to point in the desired direction. A cosine taper in both planes was assumed for the overall aperture, and 25 degrees Kelvin was assumed as a normal receiver temperature. The aperture distribution of the array was taken to have the form

$$F(x, y) = \cos(\pi\sqrt{A}\, x) \cos(\pi\sqrt{A}\, y) \tag{29}$$

in which the effects of the illumination discontinuities at the subaperture edges have been neglected.

The corresponding gain function for the array is then

$$G(u - u_o, v - v_o) \approx \eta \frac{4\pi A}{\lambda^2} \frac{\pi^8}{16^2} \frac{\left[\cos \frac{\pi\sqrt{A}}{\lambda}(u - u_o)\right]^2}{\left[(\frac{\pi}{2})^2 - (\frac{\pi\sqrt{A}}{\lambda})^2 (u-u_o)^2\right]^2} \frac{\left[\cos \frac{\pi\sqrt{A}}{\lambda}(v - v_o)\right]^2}{\left[(\frac{\pi}{2})^2 - (\frac{\pi\sqrt{A}}{\lambda})^2 (v - v_o)^2\right]^2} \tag{30}$$

where

$$\eta = 0.656$$

$$\frac{\sqrt{A}}{\lambda} = 4673$$

The numerical evaluation of the temperature integral over the sun for this size array would take a prohibitive time on the GE time-sharing computer, so the integral was evaluated analytically by approximating the sun by a circumscribed square. Thus the temperature computed by this method is slightly higher than the true value. The integration for the temperature yields the following result

$$T_{es}^{a} < T_s (0.656) \frac{\pi^3}{128} I_1 (\hat{x}_1, \hat{x}_2) \tag{31}$$

where

$$I_1(\hat{x}_1, \hat{x}_2) \approx \left\{ \pi - \frac{1}{(\hat{x}_2 + \pi)} - \frac{1}{(\hat{x}_2 - \pi)} - \mathrm{Si}(\hat{x}_1 + \pi) + \frac{1 - \cos(\hat{x}_1 + \pi)}{\hat{x}_1 + \pi} \right.$$

$$\left. + \frac{2}{\hat{x}_2} - \mathrm{Si}(\hat{x}_1 - \pi) + \frac{1 - \cos(\hat{x}_1 - \pi)}{(\hat{x}_1 - \pi)} - \frac{1}{\pi} \left[\mathrm{Cin}(\hat{x}_1 + \pi) - \mathrm{Cin}(\hat{x}_1 - \pi) \right] \right\} \tag{32}$$

where

$$\mathrm{Si}(x) \equiv \int_0^x \frac{\sin t}{t}\, dt \tag{33}$$

$$\mathrm{Cin}(x) \equiv \int_0^x \frac{(1 - \cos t)}{t}\, dt \tag{34}$$

$$\hat{x}_1 \equiv \frac{2\pi\sqrt{A}}{\lambda} (u_1 - u_o)$$

$$\hat{x}_2 \equiv \frac{2\pi\sqrt{A}}{\lambda} (u_2 - u_o)$$

where $u_2 = \sin 0.25° = -u_1 = 0.00436$

The upper bound on T_{es}^{a} given by Equation (31) is plotted in Figure 4-28.

When the calculated temperatures are put into Equations (24) and (27), the curves shown in Figure 4-29 are obtained. Here, the effective array temperature, T_{es}^{a}, is plotted as a function of the angular separation between antenna pointing direction and edge of the sun for zero amplitude errors and for rms amplitude errors of 0.10 and 0.20 db. It is evident that over the range shown, the zero error array temperature contributes a negligible amount to the effective temperature; and that for the cases shown, the array cannot be steered closer than one sun diameter from the edge of the sun, if the increase in temperature due to the sun is not to exceed 25°K.

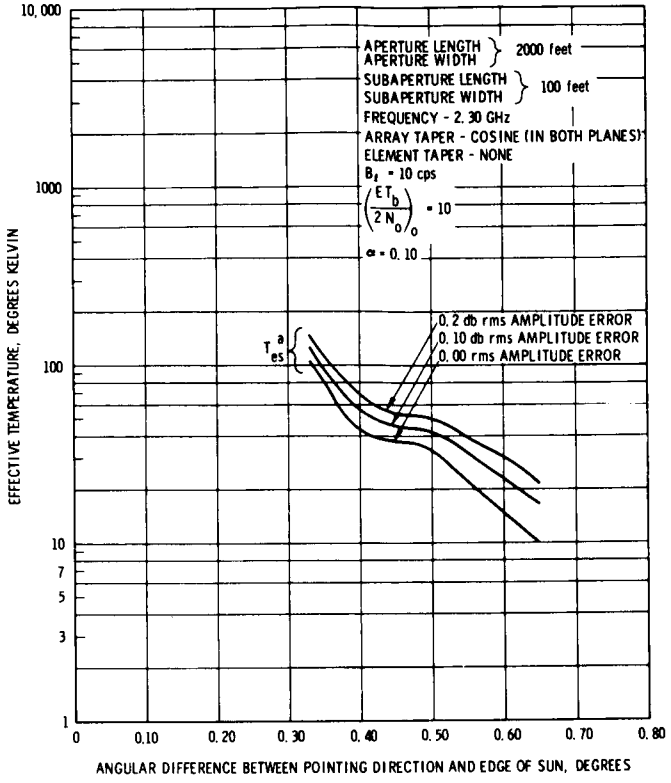


Figure 4-29. Effective temperatures with errors.

As stated in Equation (22), the gain of the array is also decreased so that the signal-to-noise ratio in the communication channel is further decreased. For the case considered, the effective reduction in array mean gain was also computed as the array was steered near the sun. The change in gain was found to be much less serious than the increase in noise temperature (see Figure 4-30). It was also found that maximum reduction in gain of the mean pattern is about 0.4 db for the 25°K temperature rise considered. This value of gain reduction includes the effects of a 0.20 db rms amplitude error. For zero amplitude errors, the gain reduction was about 0.25 db at the same pointing angle.

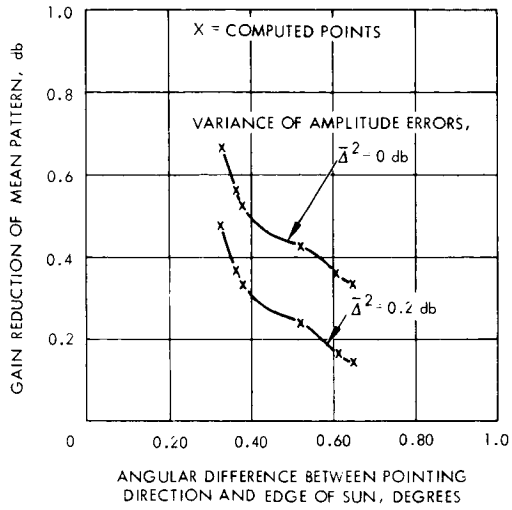


Figure 4-30. Gain reduction.

Summary. The effect on antenna noise temperature of steering an adaptive array near the sun has been considered. The analysis has been carried out under the assumption that the noise due to the sun in the phase-locked loops is statistically independent although in actual practice there will be some correlation from element to element. To assess the effect of the sun noise accurately, the cross-correlation of the noise from element to element should be taken into account. The results presented here, and based on an assumption of independent noise tend to be optimistic.

Phase and Time Delay Compensation for Large Arrays. In the design of a very large ground antenna array for the reception of high-bit-rate data from a distant space vehicle, provision must be made for the

summation of signals from the various portions of the array. It is essential that the signals received from the various elements or subapertures of the total array be added with proper regard to RF phase if the desired gain and pattern are to be achieved. Moreover, such addition with regard to phase must be performed even though the aperture may be of such large extent that the incident wavefront can no longer be regarded as plane. Atmospheric turbulence and other inhomogeneities may be sufficient to produce a departure from a planar wavefront by an amount that is not small compared with the wavelength. A second problem, that of time delay compensation, is distinct from the RF phasing problem in that such compensation requires a true time delay correction rather than a phasing correction.

An array of n elements or subapertures is represented schematically in Figure 4-31, and a planar wavefront is indicated as being incident on the array. The plane of the wavefront is taken as being at an angle θ with respect to the plane of the aperture. The signals

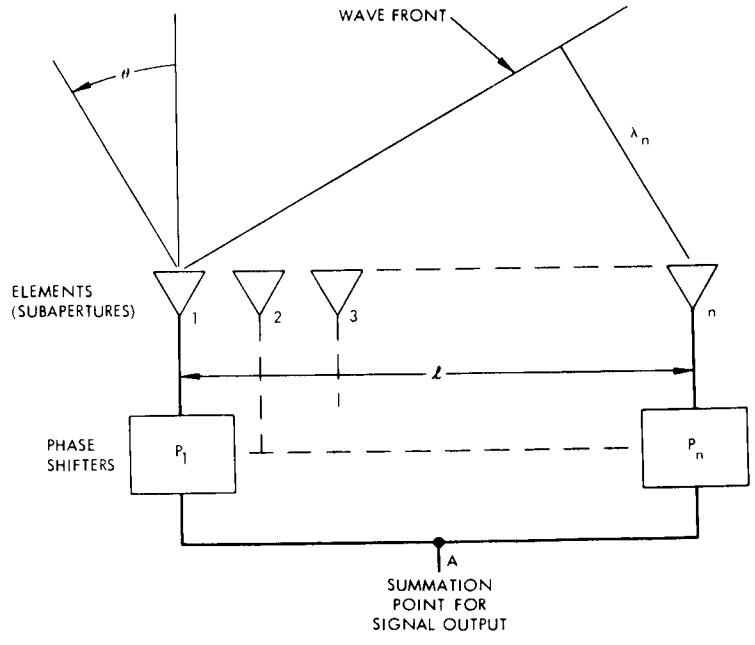


Figure 4-31. Large array schematic.

received at the n subapertures are presumed to be added at the point A to provide the output signal. In order that the maximum output signal and the maximum antenna gain be achieved at a single frequency, the RF phases from the various portions of the total aperture must be adjusted in the phase shifters $p_1 \ldots p_n$. If the phase of the signal at element number 1 is taken as the reference phase, then the phase delay in the arrival of the signal at element n will be

$$\phi_n = \frac{2 \pi X_n}{\lambda_o} \tag{35a}$$

or

$$\phi_n = \frac{2 \pi \ell \sin \theta}{\lambda_o} = \frac{2 \pi f_o \ell \sin \theta}{c} \tag{35b}$$

where

ℓ = the distance across the array between phase centers of elements 1 and n

λ_o = the wavelength corresponding to a particular single frequency, and

c = the velocity of propagation

At a single frequency this phase delay can be introduced, modulo 2π, at phase shifter p_1 to provide a signal maximum at the output. The phase shift introduced is then

$$\phi'_n = \frac{2 \pi \ell}{\lambda} \sin \theta - 2 m \pi \tag{36}$$

where

$0 < \phi_n < 2 \pi$

m = 0, 1, 2,.....

If the phase adjustments are made without concern for integral multiples of 2π radians, the total range of phase adjustment needed in the phase shifters is restricted to 2π radians or 360 degrees. Errors in the phase compensation will occur, however, that limit the bandwidth. At a frequency removed from f_o by an amount Δf, the phase shift introduced is

$$\phi_n + \Delta\phi_n = \frac{2\pi\ (f_o + \Delta f)\ \ell \sin\theta}{c} \tag{36}$$

or

$$\Delta\phi_n = \frac{2\pi\ \Delta\ f\ell}{c}\ \sin\theta \tag{37}$$

If a limitation in allowable phase error across the array of 2π radians is assigned, the bandwidth becomes

$$\Delta f = \frac{c}{\ell}\ \frac{1}{\sin\theta} \tag{38}$$

This expression provides a rule of thumb: for a large pointing angle, θ, the bandwidth of the array is limited to c/ℓ, unless real time delays corresponding to multiples 2π can be introduced.

The significance of the bandwidth limitation can be seen in the following way. If there is any difference in the times at which a signal is received at two places on the array and if the signals received at these two points are brought together by means of a suitable transmission line, the carrier phases at the transmission line outputs can be brought into coincidence by shifting the phase of one of the received carriers by a suitable amount, modulo 2π. It can then be supposed that the received signal is modulated and that the net differential time delay between the two signals, as seen at the outputs of the transmission lines, is on the order of a microsecond. If the reciprocal of the modulation bandwidth is small relative with this differential delay, the modulation output of the one transmission line at a given time will, in general, have no relationship to the modulation output of the second transmission line at the same time. In particular, the two transmission line outputs cannot be summed to improve the signal-to-noise ratio, as such a

summation process only serves to garble the received modulation. If the assumption is made that the information is in digital form, it is quite apparent that the modulation will be distorted whenever the difference in propagation time for signals to be summed is significant relative to the bit length. The condition that must be satisified if the summation of the signals of the outputs of the two transmission lines is to preserve the intelligence contained in the modulation and increase the received signal-to-noise ratio, is that before the signals are added, the differential time delays must be corrected to within a small fraction of the reciprocal of the modulation bandwidth. This same argument is completely general and applies to the case in which signals are received at any number of subapertures and brought to a common point for summation.

The effect of random errors in the settings of the phase shifters has been discussed in Section 4. 5. 2. The precision with which the phase taper across the array must be set, module 2π, to obtain a given pointing accuracy can be expressed in terms of the array beamwidth. For simplicity the aperture is taken to be circular. It is assumed that a linear phase taper exists across the aperture diameter such that the phase difference between the two end elements is ϕ. The relation between ϕ and the beam position, θ, is then given by the relation

$$\phi = \frac{2\pi d}{\lambda} \sin \theta \tag{39}$$

where

d = array diameter

λ = wavelength

θ = angle measured from broadside

If two closely adjacent beam positions, θ and θ_1, are considered, there results

$$\phi_1 - \phi = \frac{2\pi d}{\lambda} \left[\sin \theta_1 - \sin \theta\right] \tag{40}$$

When advantage is taken of the trigonometric identity,

$$\sin \alpha \sin \beta = 2 \cos \frac{1}{2} (\alpha+\beta) \cdot \sin \frac{1}{2} (\alpha-\beta)$$

and use is made of the approximations,

$$\cos\left(\frac{\theta_1 + \theta}{2}\right) \cong \cos\theta$$

$$\sin\left(\frac{\theta_1 - \theta}{2}\right) \cong \frac{\Delta\theta}{2}$$

then Equation (40) can be reduced to

$$\phi_1 - \phi = \Delta\phi \cong \frac{2\pi d}{\lambda} \cos\theta \ \Delta\theta \tag{41}$$

The 3-db beamwidth of a uniformly illuminated circular aperture of diameter d is given by Silver[24] as

$$BW = 1.02\,\frac{\lambda}{d} \cong \frac{\lambda}{d} \tag{42}$$

Since the effective aperture length for a scanned array is proportional to the cosine of the scan angle, the beamwidth as a function of the scan angle θ is given by

$$BW(\theta) = \frac{\lambda}{d\cos\theta} \tag{43}$$

From Equations (41) and (43) then,

$$\Delta\theta = BW(\theta)\,\frac{\Delta\phi}{2\pi} \tag{44}$$

In Equation (44), $\Delta\theta$ can be interpreted as the beam pointing error that results from an error in the phase taper, where the phase error at the end element caused by the error in phase taken is $\Delta\phi$. It can be seen that if the steering error is to be held to within a quarter of the 3-db beamwidth, the phase error across the array must be less than $\pi/2$ radians.

It can be assumed that within the subapertures the signals from the various elements are to be summed at the radio frequency. If the signals received from several subapertures are also to be summed at the radio frequency and if time delay between subapertures must be introduced, then to preserve pointing direction and coherence for the modulation the time delay introduced must be known to the same accuracy within some given interval of 2π radians. In other words, the sum of the

time delay and the phase shift must be held to an error of less than $\pi/2$ radian, modulo 2π, across the array where the number of integer multiples of 2π has been reduced. If the signals have been detected at the outputs of the subapertures, however, the tolerance on the time delay is reduced.

Subaperture Size. Although a precise determination of the optimum subaperture size cannot be given at this juncture, a summary of the principal factors that influence the choice of this dimension can provide a preliminary estimate. A few assumptions, of course, must be made.

A large array is represented schematically in Figure 4-32; the aperture can conveniently be considered to consist of subapertures that in turn may consist of subsections.

A subaperture is considered here as that portion of the antenna over which the signals are added linearly at the radio frequency without the introduction of a special time delay or signal processing. In the instance of an array of large parabolic reflectors, the subaperture would be the aperture of a single reflector. In the case of an electronically scanned array, the subapertures would consist of subsections or elements.

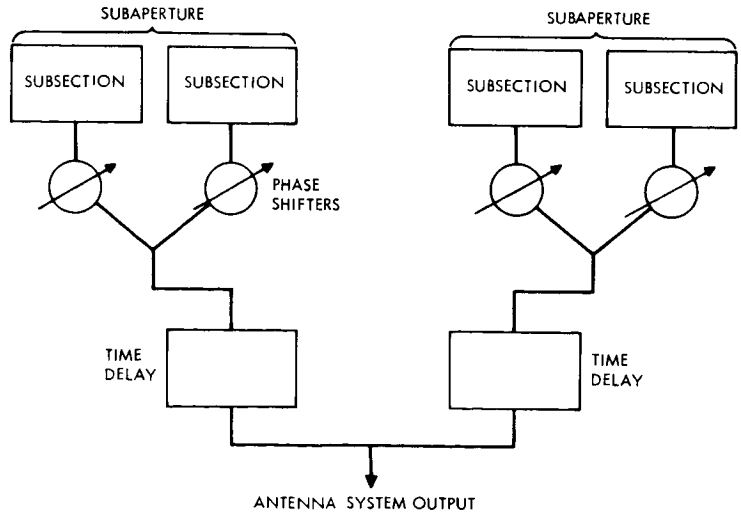


Figure 4-32. Subapertures for large array.

A very large antenna array should be usable at large scan angles which, for sake of illustration, can be taken as 60 degrees. At such large angles the signal received at the rear edge of the aperture will lag that received at the front edge by a time approaching the time required for an electromagnetic wave to propagate a distance equal to the aperture dimension. Since pure time delay compensation within a subaperture appears inappropriate on the basis of cost and complexity, it follows that the maximum dimension usable for a subaperture in an electronically scanning antenna must be small relative to the equivalent spatial length of one bit period, that is, the distance travelled by an electromagnetic wave during the bit period.

Detailed analysis will be required to determine exactly the extent of garble and modulation distortion introduced as a function of aperture dimension. This analysis is straightforward, if perhaps laborious for any particular modulation system. A reasonable choice, which would be expected to be confirmed by the details of this analysis when performed, would be an aperture dimension that did not exceed about one-tenth of the spatial equivalent of a bit period at the highest bit rate for which the receiving system is to be designed. Beyond this point the system's self noise should become noticeable, while with aperture transit times not exceeding one tenth of the bit period, the integrated effect over the aperture should be essentially negligible. If a megabit data rate is assumed, the dimension corresponding to the duration of one-tenth of a bit is approximately 100 feet.

Although the size of a mechanically gimballed reflector antenna is not limited by propagation time within the subaperture, since the plane of the aperture is aligned with the incoming wavefront, the size of a mechanically positioned array could be so limited, depending upon design, by propagation time or losses within the feed system. The size of parabolic reflector antennas tends rather to be limited by cost, because of the complexity of the structure required to achieve the required surface tolerances and also because of the precision demanded of the servo-mechanisms with which the increasingly

narrow beam is pointed. The cost rises rapidly as size is increased; reference to Figure 4-14 (estimated costs of antenna systems versus occultation time) indicates that a diameter of approximately 100 feet, although not a firm figure, appears a reasonable choice.

Two important factors affect the determination of the minimum subaperture size. As the size of the subaperture decreases, the necessary signal distribution and processing systems become more extensive and complex and hence, more expensive. The advantage from this standpoint lies in using as large a subaperture as is practicable. Second and more significant, the subaperture must be made sufficiently large to permit the acquisition by some means of a coherent signal for processing.

The matter in question is that of obtaining a coherent carrier for demodulation of the intelligence on the incoming wave at each individual subaperture. In a somewhat more general context, this problem reduces to an adjustment of the carrier phases of the signals received at each of the many subapertures so that the carrier signals can be added in phase. Many alternatives are available for the achievement of this objective.

The heart of the carrier phasing problem lies in the means employed for initial acquisition. Initial acquisition is important because, once a suitable composite carrier reference is available, the carriers received from each individual subaperture can in principle be compared with this composite and their phases adjusted accordingly. Once acquisition has been effected, closed-loop control techniques can assure that the carriers are thereafter appropriately phased relative to one another.

The first, and most straightforward, means for acquisition of the carrier is the employment of a separate phase-locked receiver for each individual subaperture. Acquisition by this technique implies that sufficient carrier power must be available in the output from the subaperture to permit lock-on; this requirement, in turn, may under some circumstances have significant implications for the division of signal power between the carrier and the information modulated onto it.

Clearly, any signal power which must remain in the carrier to facilitate initial acquisition is no longer available for information transmission. In the extreme case, situations are conceivable in which the signal strength at the output of the subaperture would be insufficient to permit carrier lock-on even if the full power of the signal resided in the carrier. It is also important to observe that, when a separate phase-locked receiver is employed at each subaperture for carrier acquisition, attention must be paid to the pattern of the individual subaperture. That is, amplitude taper over the complete array may have very beneficial effects for the array as a whole but may do nothing to control the sidelobes, and hence the received backlobe or sidelobe radiation in the output of the individual subaperture. Since the lock-on characteristics of the phase-locked receiver associated with the subaperture are influenced only by the signal-to-spectral-noise density in the subaperture output, sidelobe control to assure a low level of backlobe and sidelobe thermal radiation may well prove essential.

It was considerations of this nature that prompted the analysis in the preceding subsection of the effect of the sun's presence in the subaperture pattern on the overall noise temperature of an adaptive antenna.

Alternative techniques for the achievement of carrier lock-on are possible. In one technique that has been proposed the signals received from all subapertures are added with various phase shifts introduced before addition; the resulting set of signals formed by this summation process is then searched to find one in which the carrier components have added nearly enough in phase to permit acquisition and carrier tracking. It is implicit in this technique, of course, that all signals must be brought together by means of suitable transmission lines. Two distinct configurations are possible. In the first place, all signals may be brought together directly at radio frequency via transmission lines driven by suitable low-noise preamplifiers. For long transmission lines this arrangement would be an expensive procedure in view of the number of subapertures that would be employed in a typical large ground array.

An alternative is the distribution of a common local oscillator signal to each of the subapertures. Distribution would be by transmission lines or microwave links driven by suitable transmitters or line drivers. The local oscillator signal would be amplified as required and then mixed with the received signal at each individual subaperture; the resulting signal would then be transmitted to a common point by means of a second set of transmission lines. The losses would be very modest at intermediate frequencies, so that suitable power levels would be provided by the low noise amplifiers. In any case, however, a low-noise RF preamplifier would be required at the individual subaperture to avoid signal-to-noise degradation in the mixer.

As can be seen from the foregoing discussion, it is immaterial how or at what frequencies the signals received from the individual subapertures are transmitted to a common point so long as the signal-to-spectral-noise-density ratio is not degraded unduly and so long as, in the event that a frequency translation is employed, a local oscillator signal derived from a common source is provided to each subaperture. A recent investigation has shown that, if each signal is split into two signals whose phases differ by 180 degrees and if, for a total of n subapertures, the 2^n sums are formed by a choice of one of the two signals from each subaperture and a summation of the chosen signals in all possible ways, at least one of the resulting 2^n signals will exhibit a carrier component at least equivalent to that which would be obtained with a composite aperture 40 percent as large as the composite formed from the total of n subapertures, if perfect phasing is supposed. Half the resulting 2^n sum signals are redundant since to each of the sum signals there corresponds another that differs only in sign. It follows from this observation that, in practice, only a subset of 2^{n-1} signals would be formed. For large numbers of subapertures, say on the order of several hundred, the resulting number of combinations clearly becomes so large as to render this method of acquisition infeasible if an equivalent 40 percent efficiency is to be achieved.

An alternative would consist of the introduction of phase shifts chosen from a table of random numbers in the outputs of each of the subapertures and the addition of the resulting signals. Repeating this

process only a modest number of times would quite likely result in a carrier component sufficient for acquisition. This possibility requires further investigation. The analysis should prove entirely straightforward inasmuch as the random phase shifts introduced in each summand before summation convert the problem into the classical one of random walks. It is readily seen from this observation that the resulting amplitude of the carrier component will have a Rayleigh distribution.

It is, perhaps, important to observe that the same technique can be applied to any collection of the subapertures, rather than to the totality thereof. It may in some cases be feasible to assure carrier lock-on without the complications inherent when the signals from all the subapertures are handled simultaneously for this purpose. The phase-splitting technique described previously might prove feasible if a relatively small number of subapertures were employed. It should also be observed that either the phase-splitting or the random phase shift technique could be employed in time sequence rather than simultaneously in parallel channels if the time available for acquisition permitted. Sequential acquisition by this means would reduce the amount of equipment required and would result in very substantial cost savings. In any event, an acquisition scheme that utilizes more subapertures than are really required to provide an adequately clean carrier reference signal should be avoided since unnecessary proliferation can prove both complex and costly.

There are several additional points on the subject of acquisition that should be noted. First, if all signals are transmitted to a common point, either at radio or intermediate frequency, all phase-locked receivers can, in principle, be located at this point rather than distributed with one in the neighborhood of each subaperture, in the event that the system is configured so that each subaperture is provided with its own phase-locked receiver. This practice of having all phase-locked receivers in a common central facility has a number of desirable features, but it may lead to very severe isolation problems.

A second point is probably of more importance from a practical standpoint. Summation of the outputs of a number of the subapertures to obtain a reinforced carrier component can, up to a certain point, be

performed on the basis of a priori knowledge. That is, until the total dimension of the region scanned by a collection of subapertures becomes so large that decorrelation caused by atmospheric or ionospheric inhomogeneities becomes significant, the phase corrections to be introduced in the outputs of the individual subapertures can be determined on an a priori basis from a knowledge of the angular position of the source. The resulting phase corrections can then be made by means of variable phase shifters before addition. As a practical matter, the number of subapertures which are ganged in this fashion will be limited by the phase stability of the various circuits and transmission lines and by the accuracy with which the angular position of the source is known; the phasing problem for a collection of subapertures is equivalent to the beam-pointing problem for a single antenna whose dimension is equal to the maximum separation between the subapertures in the group employed in this way.

A final observation may be made that demodulation of the received data need not necessarily be performed at a central point. As long as a coherent carrier signal is available at each subaperture, the output of the subaperture can be mixed with a suitably phased version of the carrier reference to obtain the baseband information. The resulting baseband may still consist of a number of subcarriers or of commutated pulse-code-modulated data channels. The important point is that, if more convenient, the data can be made available in video form at the output of each individual subaperture, so that video transmission lines may be employed to bring the data together at a single point for time delay compensation and coherent addition. Such demodulated data will be available almost automatically if a separate phase-locked receiver is employed at each subaperture, but any other provision for a coherent carrier will serve as well for this purpose.

Since distribution system costs and acquisition and processing problems tend to decrease with increasing subaperture size, it would appear that the subaperture should be as large as the indicated upper

bounds permit. It is concluded with reasonable assurance that a subaperture dimension of approximately 100 feet across is appropriate to the system parameters assumed, being large enough, in general to facilitate carrier acquisition and small enough to avoid modulation distortion.

Further discussion of the subject of time delay compensation for large arrays is given in Appendix E.

Components for Ground-Based Antennas. It is advantageous at this point to review the principal requirements to be placed on the ground-based antenna system. The optimum antenna system for this application is one that maximizes the signal-to-noise ratio under the practical constraints of cost, tolerances, reliability, and noise environment.

The antenna must have a low equivalent noise temperature and must provide high gain. These requirements imply a pattern of high directivity with low sidelobes and, hopefully, low grating lobes. Further, the losses in the distribution system and in the phase shifters should be low.

The high-gain pattern must be steerable through a wide angle in the equatorial plane to provide a practical minimum of the number of sites around the earth. Relatively wide angular steering in a direction perpendicular to the equatorial plane is also required as determined by the width of the region in which spacecraft trajectories may lie and the particular locations of sites. Because of the necessarily high directivity of the antenna system, angular pointing must be extremely accurate. It is significant, however, that only very low angular tracking rates are needed.

Since the orientation of an antenna on a spacecraft can be expected to change relative to a given ground-based antenna during the course of a mission, and since provision must be made to communicate with more than one spacecraft, it is extremely desirable if not mandatory that the antenna system be circularly polarized.

As has been discussed earlier, there are basically two types of systems that can be used in a large ground-based antenna, a large mechanically steerable reflector and a phased array with a stationary or fixed aperture comprised of subapertures. The subapertures of

the latter system consist of radiating elements whose relative phasing controls the pointing direction of the antenna beam.

The steerable paraboloidal reflector has been shown to be economically and technically practical for antenna aperture sizes on the order of a few hundred feet. For aperture sizes on the order of a thousand feet, however, a single steerable paraboloid does not appear feasible today and has very little probability of being practical in the next ten or twenty years.

The mechanical problems that limit the size of gimballed antennas do not limit the size of the array. Rather the problem of time-delay compensation becomes significant. There are also at present practical problems of cost, reliability, and maintainability that must be solved before the thousand-foot array becomes feasible.

A phased array consists of radiating elements, a power distribution or collection network, a beam-steering or phasing system, and an optimal number of low noise preamplifiers. Each of these antenna subsystems plays an important and interdependent role in determination of the overall antenna performance. There exist a variety of phased-array techniques: these include the use of a phase shifter at each element; the use of a multiple-beam, multiple-port antenna with switches to steer the beam; and the use of a mixing scheme that translates a phase shift from the operating frequency to a convenient frequency band. Those areas in phased-array distribution and phase-shifter technology that must be advanced to make the array practical are delineated in this subsection.

Distribution Networks. The distribution network collects the signal from each of the radiating elements and phase shifters of the array and brings them to a common receiving port so that they combine in phase with a minimum of loss. The distribution network largely and sometimes wholly determines the antenna aperture distribution; hence, it determines the antenna pattern, sidelobe level, and directivity. Distribution systems are taken here to include those in which the distribution is essentially optical. Various types of distribution networks described here are for multiple-beam systems, phase scanned systems,

and low-noise systems. Areas of further study and development beyond the present state of the art that are necessary to satisfy the present requirements are pointed out.

Multiple-Beam Distribution Systems. Several kinds of multiple-beam antenna systems have been devised. The distribution system which forms the multiple beams determines the geometrical and mechanical characteristics of the antennas. The three types briefly indicated in the following paragraphs illustrate the difficulties attendant on their application, most of which are due to circuit or geometrical size and complexity.

The use of a Butler matrix for beam forming is practical where 2^n outputs are desired. The Butler feed has the same number of inputs as feed outputs, and simultaneous beams with uniform aperture distributions may be formed at each input. Other distributions can be obtained through combinations of adjacent beams that yield cos x, $\cos^2 x$ on a pedestal, and higher order cosine distributions. The matrix configuration is excessively complex for most applications that do not specifically require multiple simultaneous beams.

Studies have been made of various spherical lens configurations which are capable of simultaneous multiple beams, true time delay, high gain, and polarization diversity. Typical of these devices is the constrained lens, in which energy, rather than passing through a dielectric lens, is constrained to controlled paths established by cables between pickup and transfer surfaces. Modifications of this basic constrained lens have also been studied which provide increased aperture utilization. One of these employs a reflector in a manner similar to the Cassegrain antenna. Another employs a dielectric correcting lens in a manner analogous to the Mangin mirror. A third version employs a dielectric hemisphere as the phasing mechanism and saves considerable cabling. Each of these lens systems provides true time delay and multiple beam capability and has advantages in packaging and efficiency over similar systems such as the Luneberg lens. All versions suffer from the disadvantage that the spherical surfaces required for true time delay phasing would result in a severe construction problem in very large arrays.

Analytical and experimental studies have been conducted with lenses that utilize two-dimensional geodesic surfaces for multiple beam forming. Extremely wide scan capability is possible with these geodesic-type lenses, which also provide simplicity, loose tolerance requirements, and broadband operation. However, the curved phasing surfaces of the geodesic configurations also imply construction complexity.

Phase-Scanned Systems. Phase-scanned antennas may be considered to fall into either of two general types. In one type, the radiating structure is fed from behind either in an optical fashion with an auxiliary feed or with a system of transmission lines or waveguides. A so-called feed-through distribution system is used, and the phase shifters required are of the transmission variety. The second general type of phase-scanned antenna consists of a radiating structure fed from the front by a primary feed. The phase control is achieved with phase shifters of the reflecting type located behind the elements of the radiating surface. This kind of antenna is ordinarily designated as a reflectarray. Since both antennas are scanned, contiguous apertures can be used.

Feed-through arrays utilizing a transmission line network for the feed offer the advantage that the aperture distribution can be closely controlled. Corporate feeds provide convenient power division where a uniform distribution to 2^n outputs is desired. Tapered distributions for low sidelobes or output configurations that require unequal power division in the tees are more complicated to achieve but are still practical. The corporate feed may also be used for monopulse; however, the step amplitude discontinuity between each half of the aperture results in sidelobe degradation on the difference pattern. The disadvantage of parallel feeding techniques such as the corporate feed at the radio frequency for large arrays is the great amount of transmission line required and the resultant loss.

In the front-fed reflectarray, a planar array of radiating elements is utilized as a phase reflecting surface in a manner analogous to the operation of a parabolic reflector. Radiation from the feed horn is collimated by the phase-reflecting surface to form a beam. The beam is steered by adjustments in the phase shift distributions of the phase shifters located at the terminations of the radiating elements. The phase shifters can be of any design that meets the other requirements such as loss and reliability.

When the reflectarray is viewed as a distribution network, it can be considered as a junction with the input terminal of the feed horn as the sum arm and all the radiating elements as output arms. In contrast to the transmission line distribution network, in which power is divided in enclosed waveguides and junctions, the input-output junction of the reflectarray is simply free space. This fact offers two advantages for application to very large arrays. Since the distribution network is basically free space, there is practically no loss due to dissipation. The transmission loss is mainly spillover loss that can be made small by proper feed design which also permits some control of the aperture illumination. The simplicity, cost, and weight of the reflectarray distribution network are better than the same characteristics in the transmission line network.

Low Noise Feed Systems. The desired aperture distribution for optically-fed arrays such as lenses, reflectarrays, etc., can be achieved with a shaped feed. An aperture distribution close to a 35-db Taylor distribution can be realized by a simple, conical, or pyramidal horn. A wide range of adjustment of amplitude taper is possible by variations in the size of feed horn and the f/D ratio. Flexibility in aperture design is possible with multiple elements in the feed. Studies have been carried out on the general requirements of feed systems to provide low noise and highly efficient operation of antenna systems. These results have been applied to the actual design and development of low-noise feed systems for front-fed and Cassegrain antennas. Among these developments are multimode shaped-beam feeds,

quasi-array feeds, and shaped-sub-reflector systems, all of which are directly applicable to reflectarray and lens-feed configurations.

A shaped primary pattern such as can be obtained from a multimode feed can minimize feed spillover loss. Such a feed may be either rectangular or circular in cross section depending on the shape of the aperture to be illuminated. Shaping to obtain a variety of aperture illuminations can be accomplished with relatively few modes in a feed. Three separate, multimode feed systems that have been developed at the Hughes Aircraft Company are representative of what can be done. These feeds provide dual-circular polarization capability, dual-plane monopulse, and dual-frequency operation. The first system was a multimode, monopulse, prime-focus feed developed to meet a high-efficiency, high-power requirement. From this system, two Cassegrain feed systems were developed; the final system is a multimode, shaped-beam Cassegrain feed used on 85-foot antenna systems. This feed provides an antenna efficiency of about 60 percent with a noise temperature of about 15 degrees Kelvin. These quantities include the losses in the monopulse bridge. At the same time, monopulse tracking to about 1/100 of a beamwidth accuracy is provided in two planes.

Monopulse feed systems for Cassegrain antennas have been studied that include four, five, eight, and twelve-horn feeds. A four-horn feed using an "egg-crate" lens structure to improve the aperture distribution has been developed for use in a 60-foot Cassegrain antenna. This feed is an S-band, dual-polarized, dual-plane monopulse feed. A five-horn monopulse feed has been designed to provide monopulse tracking at one frequency and simple listening or transmitting at a second frequency. This feed is particularly suited to a 3:2 frequency separation. A four-horn feed using a multimode shaping aperture has also been developed. This feed provides a higher aperture efficiency and lower noise temperature than a standard four-horn feed.

Investigation of the optimum shape of the sub-reflector in a dual-reflector antenna such as a Cassegrain system has also been a topic of considerable investigation. Modification of the hyperboloid edge is one example of a useful technique. The shaping of the reflector surface may also be controlled to provide particular aperture distributions with various feed systems. This technique gives an additional parameter in the control of low-noise antenna designs.

Phase Shifters. Beam steering for a deep-space communication phased array of the conventional type requires some type of phase-shifting device. The primary requirements for such a device are that it be capable of 360 degrees of phase shift and that it have an extremely low insertion loss. From system noise considerations, it would be preferable that this insertion loss be less than 0.1 db. In addition to these primary considerations, the device must be relatively inexpensive, be capable of being packaged to fit within the array element spacing, and be temperature insensitive to ambient environments. At present, there is no phase shifting device that will meet all of these requirements. As a basis for a discussion of these devices, a comparison was made of the performance parameters of present state-of-the-art devices at an operating frequency of 2.3 Ghz. The results are given in Table 4-15. As may be seen from the Table, many of the devices have high insertion losses, partly due to the usual requirements of fast switching speed and high power handling, but if the antenna is used in reception only, neither high-speed nor high-power capability is necessary. In Table 4-14 the devices were classified into five general groups. These groups, their advantages and disadvantages, and the work that would be required to adapt them to a very large phased array are discussed below.

Diode Phase Shifters. Diode phase shifters are small, lightweight units, insensitive to temperature, which can be switched from one phase setting to another in a few nanoseconds. They are, at present, somewhat costly due to the cost of the diodes and their

Type of Phase Shifter	Insertion Loss for 360° of Phase Shift in db.	Reciprocal (R) or Non-Reciprocal (NR)	Transmission (T) or Reflection (R)	Analog (A) or Digital (D)	Switching Speed	Drive	Temperature Sensitivity	Weight	Size	Cost
I. Diode Type										
1. Analog Varactor	1.2	R	T	A	Fast	Low	Low	Low	Small	Moderate
2. Digital Diode-Transmission Type	0.8	R	T	D	Fast	Low	Low	Low	Small	Moderate
3. Digital Diode - Reflect Type	0.6	R	R	D	Fast	Low	Low	Very Low	Small	Low
II. Ferrite Type										
1. Reggia-Spencer (non-latching)	0.8	R	T	A	Moderate	Moderate	Moderate	Moderate	Moderate	Moderate
2. Digital - Twin Slab	0.6	NR	T	D	Fast	Very Low	Low	Low	Small	Low
3. Analog - Twin Slab	0.6	NR	T	A	Fast	Moderate	Low	Low	Small	Low
4. Analog - Reflect	0.6	R	R	A	Moderate	Moderate	Low	Low	Small	Low
5. Reggia-Spencer (Latching)*	2.6	R	T	A	Fast	Moderate	Moderate	Moderate	Small	?
6. Digital Reciprocal Latching*	2.0	R	T	D	Fast	Moderate	Moderate	Low	Small	?
III. Ferroelectric										
1. Analog*	4.0	R	T	A	Moderate	High	Poor	Moderate	Moderate	?
2. Digital*	1.8	R	T	D	Moderate	Moderate	Poor	Moderate	Moderate	?
IV. Plasma										
1. Analog*	0.5	R	R	A	Moderate	High	Low	Moderate	Moderate	?
2. Digital*	0.5	R	R	D	Moderate	High	Low	Moderate	Moderate	?
V. Mechanical										
1. Digital Reflect	0.1	R	R	D	Slow	Moderate	nil	Low	Moderate	Low
2. Analog Reflect	0.1	R	R	A	Slow	Moderate	nil	Low	Moderate	Low
3. Digital Transmission	0.1	R	T	D	Slow	Moderate	nil	Moderate	Moderate	Moderate
4. Analog Transmission	0.1	R	T	A	Slow	Moderate	nil	Moderate	Moderate	Moderate

*Still in experimental laboratory stages

Table 4-14. Phase shifter characteristics.

mounting structure, and they have an insertion loss higher than desired for this application. A picture of a typical 4-bit, S-band, diode phase shifter is shown in Figure 4-33.

Ferrite Phase Shifters. Ferrite phase shifters are typically waveguide size, moderate in weight, somewhat temperature sensitive, can be switched from one phase setting to another in a few microseconds and require significant drive energy. They are somewhat costly, due to the cost of the ferrite material, and also have an insertion loss higher than desired for this application. A picture of a typical, four-bit, C-band, ferrite phase shifter is given in Figure 4-34.

Ferroelectric Phase Shifters. Ferroelectric phase shifters are quite small and lightweight. They are, at present, extremely temperature sensitive, due to the sensitivity of the ferroelectric crystal, and they have very high insertion loss characteristics. Since they are still in the experimental stages, production costs are unknown. At present, it appears that a major improvement will be required in the basic crystal before these devices can be considered for use in a large array.

Plasma Phase Shifters. As in the case of the ferroelectric phase shifter, the plasma phase shifter is still in the experimental stage. It is moderate in size and weight with a negligible temperature sensitivity. The insertion loss is somewhat higher than desired but a significant reduction may be possible. At the present time, it is not a low cost device and requires significant drive energy; both factors are due to the need for the generating and sustaining of a plasma.

Mechanical Phase Shifters. Because of the relatively slow scan rate requirements of the present application, mechanical phase shifters are appropriate. In fact, from an insertion loss and cost standpoint, they appear to be the most promising units. In particular, four different devices show merit. One is a reflecting type, two of which, when used with a hybrid coupler, make a precision transmission phase shifter. The remaining three devices are transmission types.

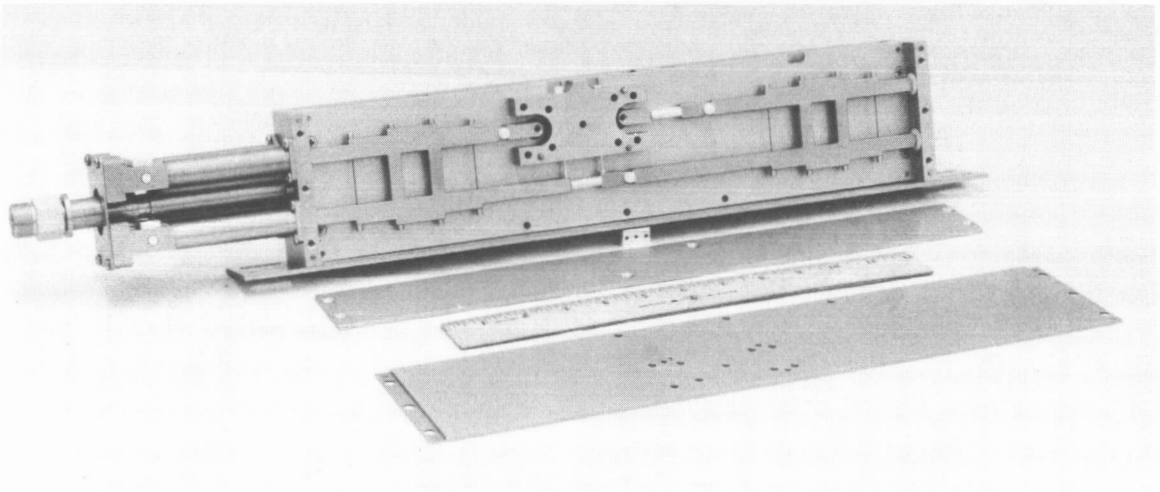


Figure 4-33. A typical 4-bit, S-band diode phase shifter.
(HAC Photo No. GS64-04-268)

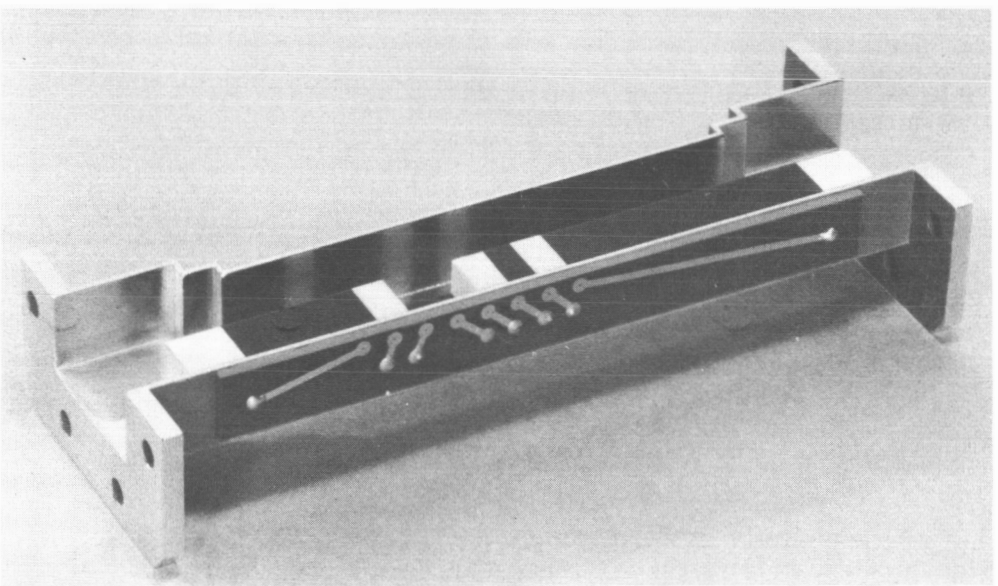


Figure 4-34. A typical 4-bit, C-band ferrite phase shifter.
(HAC Photo No. GS65-05-128)

A reflecting type of mechanical phase shifter can be built using a section of waveguide with a movable shorting plate. A sketch of such a device is shown in Figure 4-35. This relatively simple microwave design can be made to provide 360 degrees of phase shift with extremely low insertion loss. A drive mechanism to move the shorting plate that is reliable, inexpensive, and which will operate with fairly low drive power appears feasible.

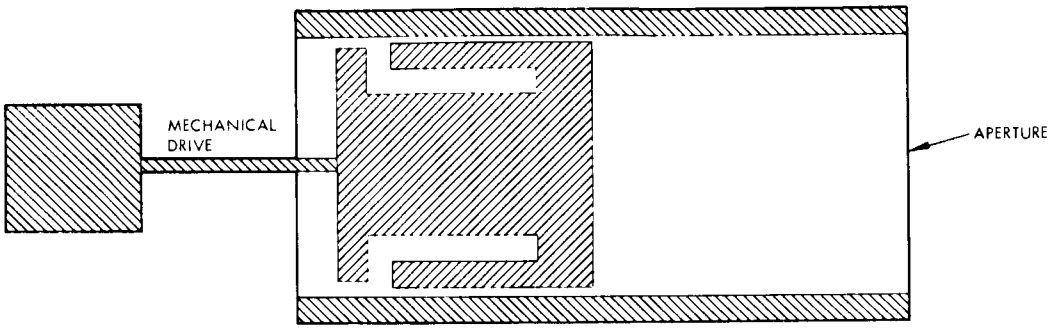


Figure 4-35. A reflecting type of mechanical phase shifter.

The first type of transmission, mechanical phase shifter uses a filter network that allows an inductive or capacitive element to be mechanically movable. Such a device has been built at S-band where a thin, stainless-steel, waveguide wall provided a variable capacitive element. The maximum wall movement was 0.025 inch and the device was capable of 360 degrees of phase shift with an insertion loss of 0.2 db. Because of the thin, stainless-steel, waveguide wall, and the small movement, the drive power was fairly low. The device was rather complex, however, and efforts would be needed to simplify the unit to reduce production costs.

The second type of transmission, mechanical phase shifter uses a movable, slow-wave structure located within a waveguide. Such a device has been built at S-band and has been used in quantities for line-length trimming. The slow-wave structure was made of aluminum to provide low insertion loss and allow die casting to be used to reduce the cost. The device was capable of 180 degrees of phase shift and had

an insertion loss of less than 0.1 db. To utilize a device of this type in a large ground-based array, efforts must be expended towards the development of a simplified, low cost, reliable drive mechanism. Possibilities of controlling the device by thermal means, air pressure, or hydraulic pressure should be explored.

A third type of transmission, mechanical phase shifter; the variable-width waveguide, uses a section of rectangular waveguide in which the width can be varied by some mechanical means. Such a device is shown in Figure 4-36. As may be seen from the figure, the device is relatively simple and should be capable of providing the required phase shift with extremely low insertion loss. As in the case of the other mechanical phase shifters, the major task in designing such a device will be to provide the proper drive mechanism.

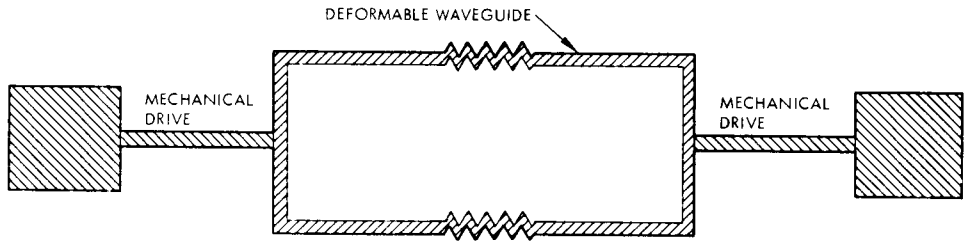


Figure 4-36. A deformable waveguide mechanical phase shifter.

At the present time, there is no phase shifter design that fully meets the requirements immediately. Many electronic phase shifters have response times much faster than required for the present application; thus there is a possibility of improvement in other parameters such as RF loss and switching power by a sacrifice in switching time. In the cases of diode, ferrite, and plasma shifters, further investigation should be carried out in anticipation of future lower costs and improvements in diodes and ferrite materials.

Because of the inherent low loss of mechanical phase shifters, these devices meet the loss requirement for a large array, and simple, reliable, low-cost drivers and driving circuits appear feasible. Consequently, mechanical phase shifters appear quite promising.

Conclusions. The reflectarray with a shaped-beam feed system appears to be one of the more promising configurations for this large array application. The placement of subapertures of this kind, side by side, permits a reasonably smooth aperture distribution to be achieved and the problem of grating lobes to be minimized. Circular polarization can also be achieved with the feed system. For a subaperture 100 feet in diameter, the feed horn would be suspended 25 to 50 feet in height. Mechanically, this arrangement is perfectly feasible, but in the event that lower feed horn height is desirable, an auxiliary reflector similar to that of a Cassegrain system can be introduced to reduce the height. A mechanical phase shifter of the reflecting type provides quite low losses and could be made inexpensively, since tolerances are quite loose. The combination of low-loss phase shifters and an optical distribution system would permit an extremely high efficiency. Although the subject of phase shifter controls is not treated in the preceding discussion, the technique of scan control using digital drivers and digital computation for the positioning of the phase shifters is currently well understood.

It is believed that at this juncture, intensive design investigations of the application of the reflectarray technique should be initiated.

4.5.6 Spacecraft Antennas

Most space vehicles are now equipped with antennas of relatively simple design, such as parabolic reflectors and feeds or arrays of slots or dipoles. Recently small electronically and mechanically despun antennas have been demonstrated. However, the size and weight of spacecraft antennas is still limited by constraints imposed by the booster vehicle and its shroud. An increase in antenna aperture would assist immensely in the improvement of the system performance characteristics. In Section 4.5.1, it is indicated that an increase of gain of about 10 db over that planned for the Voyager antenna for 1971 can be taken as a goal. The gain of this Voyager antenna is expected to be 32 db. This figure presumably represents the anticipated state-of-the-art for conventional antennas by 1971. There are limitations in several principal areas when increases in the gain are contemplated by conventional techniques which increase the aperture area:

At a given frequency and a specified efficiency, the same tolerances on the surface of a reflector must be maintained regardless of the size. Holding the tolerances becomes more difficult with increasing size and, although not impossible for ground-based antennas, would result in increased weight for a spacecraft antenna.

A second limitation is the space available in the launch vehicle for the stowing of large antennas. A single large antenna may require a shroud that is much too large or awkward for the vehicle to handle. Techniques for overcoming this problem by fabrication of the antenna in segments have been studied. The antenna would then be automatically assembled after the craft left the earth's atmosphere. The application of these techniques, as well as of those that employ inflatable structures, involves difficulties in attainment of the required antenna tolerances. There is also a serious question about the reliability of some of the schemes. As larger boosters come into general use, the limitation of available shroud space will become less severe.

A third limitation on antenna size results from the constraints on steering accuracy and on precise knowledge of spacecraft orientation. If the antenna is pre-programmed to point toward the earth, then the accuracy with which the spacecraft orientation is known is significant. If the antenna has no steering capability with respect to the spacecraft, then the spacecraft orientation must be precisely maintained because of the narrow beam of a high gain antenna. The maintaining of a precise orientation may entail the consumption of large quantities of propellant over the life of a long mission.

The use of a self-steering array on the spacecraft may overcome many of the problems outlined and allow a sizeable increase in gain over the 32 db planned for Voyager. The self-steering technique would allow a large antenna to be created from numerous radiating units, each of which would have a relatively low gain and a beam sufficiently broad so as not to impose severe tolerances on spacecraft attitude control or antenna steering. The phasing of elements of a self-steering array is accomplished by the associated electronics so that the tolerance in the positions of the individual elements relative to the plane of an incident wavefront becomes unimportant, as a rule. This tolerance relaxation should minimize positioning and alignment problems in deploying or unfurling antennas.

A general discussion of important considerations in the design of self-steering arrays for spacecraft is given here. Interelement spacings and angles of coverage, a typical design of a spacecraft array, and phase and amplitude errors are significant matters discussed. Some possible mechanizations for self-steering spacecraft arrays are covered. Performances are described for arrays with separated up and down link frequencies and for arrays whose performance does not depend on such separation. Discussion of promising circuits for doppler correction is included.

Self-Steering Array Interelement Spacing. In a self-steering array it is desirable that the number of steering modules required for a given

gain be minimized. The weight, size, and cost are proportional to the number of modules and are nearly independent of aperture size and frequency.

A self-steered antenna that utilizes a conventional Butler matrix array, a self-phased array, or an adaptive array has an interelement spacing of approximately $0.5\lambda_o$. With this spacing the beams can be generated anywhere within visible space without the presence of grating lobes. The number of beams is equal to the number of elements in the Butler matrix array. In some applications it is desirable that most of the space be filled to obtain wide angle coverage. For communication probes, however, a much smaller angle of coverage is required. This limited angle of coverage permits the use of a larger interelement spacing and an element that has directivity. In this manner a higher gain array is achieved from a limited number of modules or from a matrix of a given number of terminals.

In an alternative technique for the achievement of increased gain, the interelement spacing is maintained at $0.5\lambda_o$, but the number of elements is increased. This procedure yields the best array with few compromises in performance for a small angle of coverage. However, the number of elements is quite large and the complexity of the system soon becomes prohibitive. For example, an array with a 35 db gain, after nominal losses have been accounted for, would require a $35\text{-by-}35\lambda_o$ aperture. With an interelement spacing of $0.5\lambda_o$, the array would have approximately 5000 elements, and a 70-by-70 Butler matrix, or comparable matrix of element feeds would be required. Not only is a matrix of this size beyond the state-of-the-art, but the internal losses would probably reduce the increased gain to a fraction of that corresponding to the theoretical aperture increase. Of necessity, then, larger interelement spacings have to be used.

The use of larger interelement spacings raises some problems and necessitates a compromise in the array design. The severity of the compromise depends on a number of parameters in the system such as the gain required, the coverage angle required, and the largest number of modules or the largest Butler matrix which is within

the state-of-the-art and has reasonable loss. The main compromise results from the fact that the array generates grating lobes whenever the interelement spacing exceeds $0.5\lambda_o$.

Grating lobes are actually new beams that arise to fill the space left when increased interelement spacing causes the beams from a given array to become narrower and the primary set of beams formed by the array or matrix to decrease in coverage angle. In the case of a Butler matrix, each new beam or grating lobe shares a terminal with one of the original beams at the bottom of the matrix, as shown in Figure 4-37. These grating lobes decrease the gain of the desired beams and must be suppressed as completely as possible. Hence the gain of the element must be increased and its beamwidth reduced as much as possible without cutting down on the gain of the outermost desired beams. The ideal element pattern would be a sector beam which would cover all the desired beams and then abruptly drop to zero to suppress all the grating lobes. However, it takes a very large aperture for the element to create

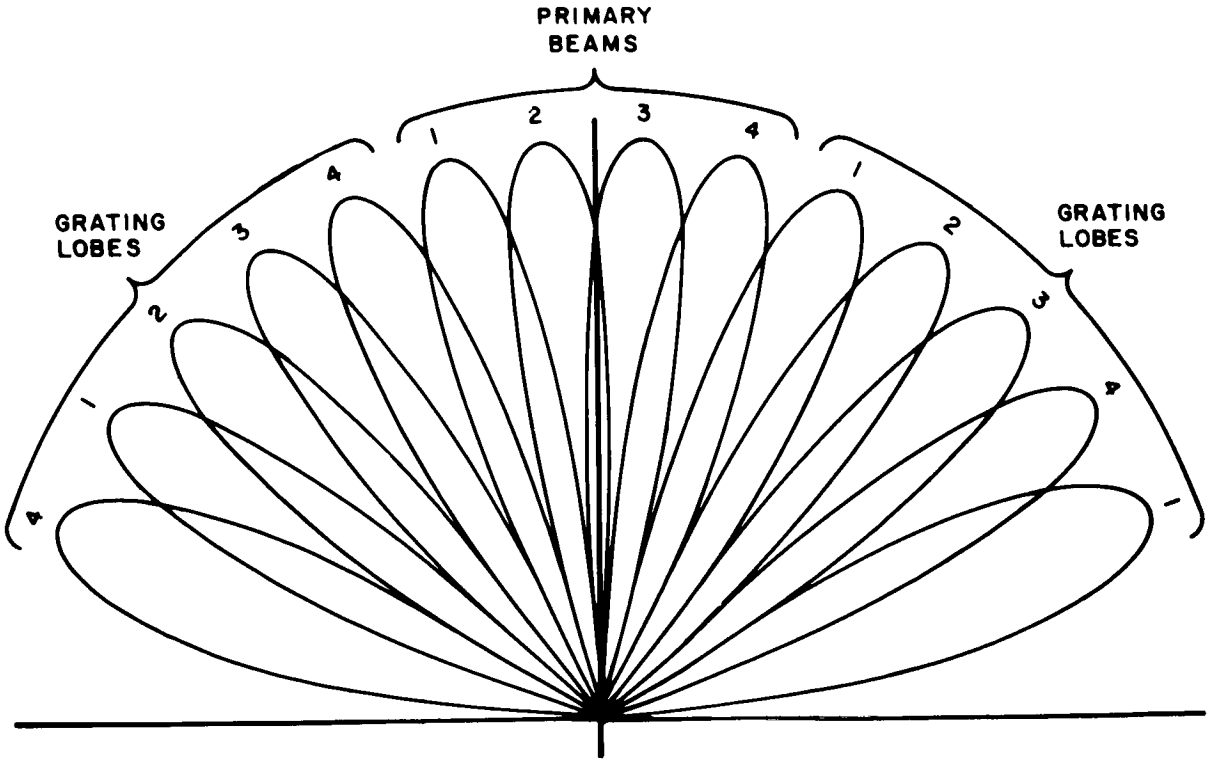


Figure 4-37. Generation of grating lobes by Butler matrix that uses interelement spacing considerably larger than $\lambda_o/2$.

even a poor approximation to a sector beam, much more than the aperture available to each element even though the elements have been spread apart considerably.

There is a straightforward solution to the problem of the adjustment of the element gain: The portion of the aperture available to each element can be assumed to be uniformly illuminated and the element beamwidth can be determined from the applicable equation. It can then be shown that the 3 db point of the element pattern always coincides with the outside 3 db point of the outermost principal beam. Thus, the outside beams will be reduced in amplitude by 3 db compared with those in the center of the cluster. With the assumption that the element pattern is essentially that of a uniformly illuminated segment of the aperture, the pattern would have the form

$$\frac{\sin\left(\frac{kL}{2}\sin\theta\right)}{\frac{kL}{2}\sin\theta}$$

where

k = the free space propagation constant

L = the edge length of the element area

θ = the angle off broadside in the principal plane.

At $\theta = \theta_o$ = the maximum scan angle, this pattern must be down 3 db from its maximum value.

In Table 4-15 are listed the number of modules for various coverage angles with a minimum array gain of 42 db. The data given can also be interpreted in terms of the allowable misalignment of the module with respect to an average array surface. Thus ±5° coverage angle means that the elements could be tilted ±5° with respect to the average array surface. No allowance is then made for scanning the beam itself to take into account the vehicle orientation with respect to the earth. In practice, a combination of scanning angle plus a misalignment tolerance will give the required coverage angle per module.

Coverage Angle, degrees	Number of Modules	Maximum Gain Per Module, db
±5	126	24
±10	500	18
±15	1000	15

Table 4-15. Number of modules as a function of the coverage angle for a minimum gain of 42 db (maximum gain = 45 db).

Example of Self-Steering Array Design. An 8-by-8 element planar array with uniform interelement spacing is presented in illustration of the design considerations for a self-steering array. The considerations presented in this subsection are those that were preliminary to the design of the antenna system described in Appendix D. For this type of array, with each element connected to an independent matched generator or load, the total gain in the beam-pointing direction is N times the effective gain, g, of a typical element, where g is measured in the presence of all other elements terminated in matched loads and the number of elements is N. The number N is assumed to be sufficiently large so that most elements see similar environments.

Each element radiates 1/N of the total power P. Therefore, the available power per element, p, is

$$p = \frac{P_a}{N} \tag{4-45}$$

and the effective radiated power (ERP) is

$$ERP = P_a G = N^2 gp \tag{4-46}$$

The required coverage region is a cone of half angle 15 degrees. The element factor must cover this region. To suppress grating lobes for the scanned beams, the element factor must also be small outside this region. It was therefore assumed that the 3 db points of the element factor fall at ±15 degrees. As has been indicated above, the

element pattern is assumed to be essentially that of a uniformly illuminated segment of the aperture. At $\theta = \theta_o = 15°$, this pattern must be down 3 db from its maximum value. This requires that L/λ be 1.714. The area gain is then approximately

$$g_a = 4\pi\left(\frac{L}{\lambda}\right)^2 = 36.9 \tag{4-47}$$

To account for losses due to spacing, coupling, and other factors, it is assumed that actual element gain is 1 db less so that the element gain is taken as 29.2 at broadside. It will be one half that value at 15 degrees from broadside in the principal plane, which is 14.6 or 11.6 db.

The gain is then taken as

G_{max} = (64) (29.2) or 32.8 db at broadside

and

G_{min} = (64) (14.6) or 29.8 db at 15 degrees from broadside (4-48)

Because of the periodic structure of an 8-by-8 element planar array and the large interelement spacing required to realize the desired gain, grating lobes exist. The element factor must fall off sharply, about 3 db, at the edge of the coverage angle, to suppress these lobes. To eliminate the grating lobes, it would be desirable to have the elements arranged in a nonuniform fashion that would avoid the periodic structure inherent in an 8-by-8 array. If the grating lobes can be avoided in the array factor, the element factor may be broadened somewhat; this increase will reduce the peak gain of the array but will not decrease the minimum gain for the beam scanned ±15 degrees from broadside.

The achievement of high gain for this antenna requires that the radiation pattern of each element should be sufficiently directive, and consequently the separation between the elements must be large enough, so that the effective aperture areas of the elements do not overlap. The directivity of each element is limited by the scanning requirement. The ideal element factor is of the form

$$E(\theta, \phi) = \begin{cases} \sec\theta & 0° \le \theta \le 15° \\ 0 & 15° < \theta \le 90° \end{cases}$$

so that the antenna gain remains essentially constant as the beam scans off broadside and the power radiated in the undesired regions is minimized.

A promising design approach is the arrangement of the elements in concentric circles with every element approximately the same distance from its neighbors. This circular arrangement is illustrated by the 64-element array shown in Figure 4-38. The average separation between elements in this circular array is about 2-1/4 wavelengths.

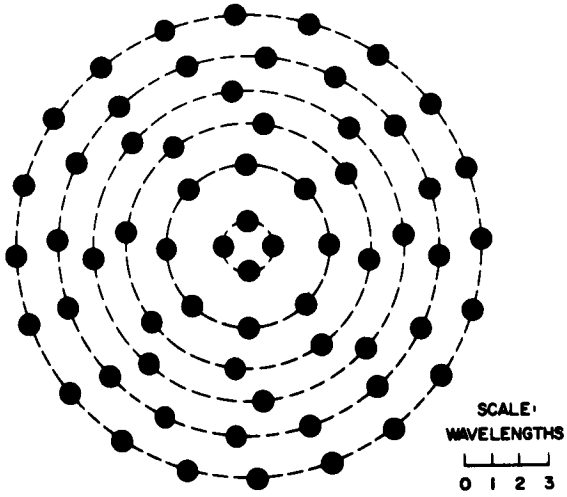


Figure 4-38. Arrangement of elements of 64-element circularly polarized array.

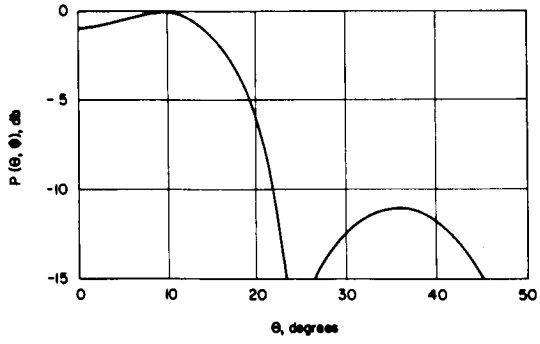
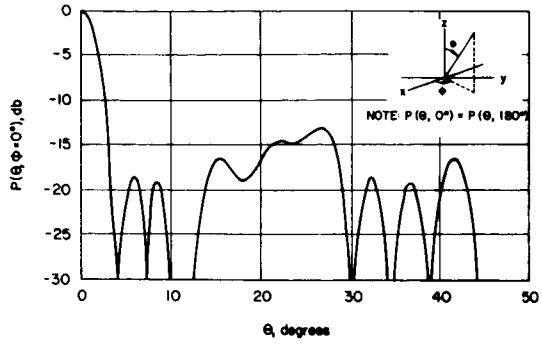


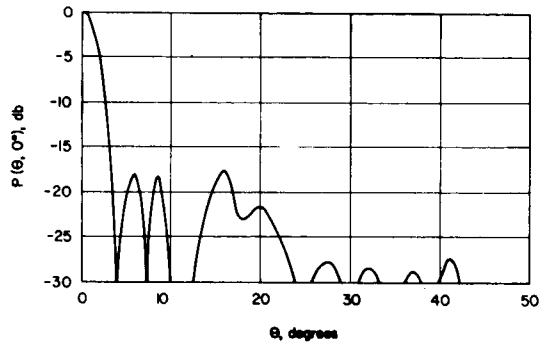
Figure 4-39. Assumed element factor of circular array. (Helical element)

Calculations were made for the radiation patterns of this array for various scan angles, with and without the assumed element factor shown in Figure 4-39. These patterns are shown in Figure 4-40. The radiation pattern of an 8-by-8 array with uniformly spaced (2-1/4 wavelengths) elements is given in Figure 4-41 for comparison; the scan angle was taken to be 15 degrees, and the element factor was included.

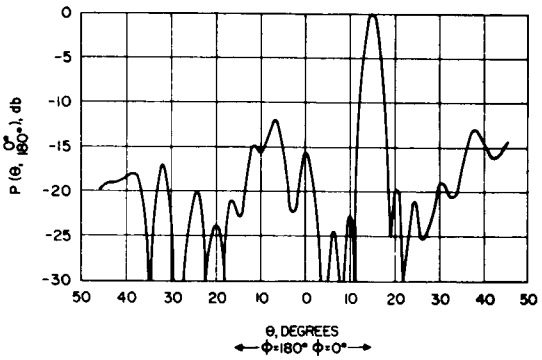
Antenna Error Considerations. Because of errors in phase and amplitude for each of the 64 elements of the circular array, the average antenna gain will be reduced, and random deviations can be expected. The first step in an analysis of such an effect is the calculation of the statistics of antenna gain with independent phase and amplitude errors assumed at each element. (The independent amplitude error assumption is



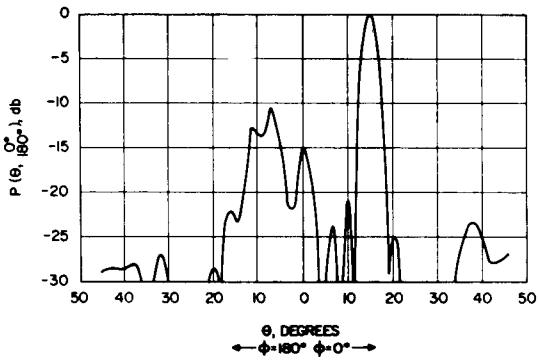
Without element factor



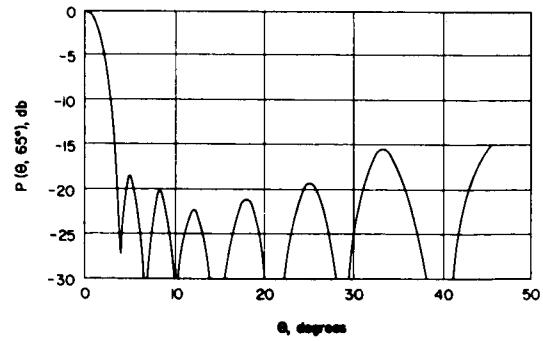
With element factor



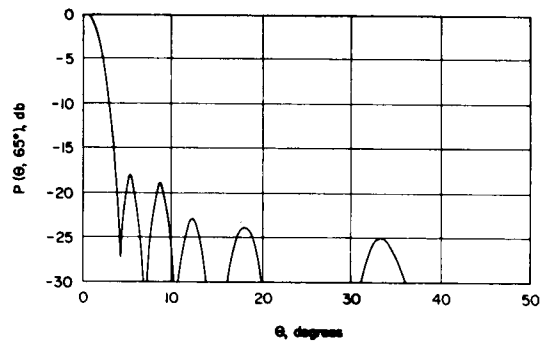
Without element factor



With element factor



Without element factor



With element factor

Figure 4-40. Calculated radiation patterns for circular array for various scan angles.

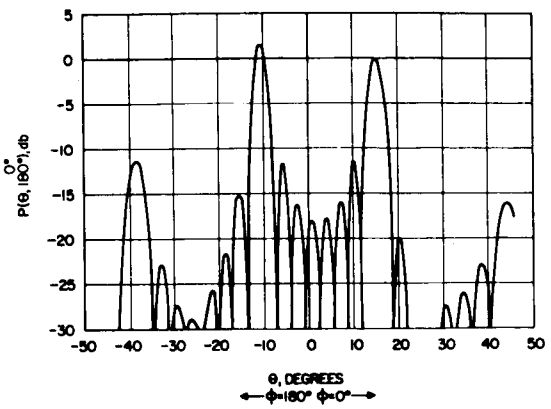


Figure 4-41. Radiation pattern of 8-by-8 planar array with uniformly spaced elements.

poor because of the usual definition of antenna gain. However, the difference will be discussed later.) For this model the gain will be defined by the ratio of the power density at broadside in the actual array to the power density at broadside in the error-free array.

From a derivation stemming from probability theory,[45] the average power gain of an error-prone antenna can be found to be

$$\overline{G} = \frac{\overline{P}(\theta_o, \phi_o)}{{}_oP(\theta_o, \phi_o)} = e^{-\sigma_\delta{}^2} \left[1 + \frac{(\sigma_\delta{}^2 + \sigma_\epsilon{}^2) \sum_{i=-N}^{N} ({}_of_i{}^2)}{\left(\sum_{i=N}^{N} {}_of_i \right)^2} \right] \tag{4-49}$$

where

$\overline{P}(\theta_o, \phi_o)$ = expected power density at broadside

${}_oP(\theta_o, \phi_o)$ = the error-free power density at broadside

σ_δ = the rms phase error for the signals at the elements

σ_ϵ = the rms amplitude error for the signals at the elements

N = the number of elements

f_i = the excitation coefficient of the i^{th} element.

For a 64-element uniformly illuminated array, there results

$$\overline{G} \approx 1 - \sigma_\delta{}^2$$

$$\sigma_G \approx \frac{\left(256\sigma_\epsilon{}^2 + \sigma_\delta{}^4\right)^{1/2}}{64} \tag{4-50}$$

$\overline{G}$ is the average gain, and σ_G is the rms deviation about this average. So, for instance, the actual antenna gain is at least $\overline{G} - \sigma_G$ with 84% probability. Inspection of Equation (4-50) shows that unless $\sigma_\epsilon^2 \ll \sigma_\delta^4$,

$$\sigma_G \approx \frac{1}{4} \sigma_\epsilon \tag{4-51}$$

The probability distribution of the random gain variability may be approximated by a χ^2 distribution. However for a high number of degrees of freedom ($\nu > 30$), the χ^2 distribution approaches a normal distribution, and, at any rate, the normal distribution percentage points will yield more conservative answers to the gain statistics questions.

The gain statistics are described by

$$\text{Prob}\left[\text{Gain} > \overline{G} - n\sigma_G\right] = P_n \tag{4-52}$$

For the normal distribution,

$P_{1.0} = 84.1$ percent

$P_{2.0} = 97.7$ percent

$P_{2.3} = 99.0$ percent

$P_{2.6} = 99.5$ percent

where the subscript indicates the multiplier for the standard deviation.

For determination of the gain and phase tolerances to be imposed upon the electronics for this system, it is desirable to have a gain degradation as small as is reasonable with a high confidence level. As an example, the gain degradation that will be allowed for this system is less than 1/2 db with a 99 percent confidence.

The 99 percent gain level is given by

$$1 - \sigma_\delta^2 - 2.3\left(\frac{\sigma_\epsilon}{4}\right)$$

For a total gain degradation less than 1/2 db from the no-error gain (with 99 percent confidence) it is required that

$$\frac{1}{2}\ \text{db} > 10 \log_{10} (1 - \sigma_\delta{}^2 - 0.575\,\sigma_\epsilon) \qquad (4\text{-}53)$$

This requirement means that any of the following combinations of rms errors would be permissible:

$\sigma_\delta < 5°$, $\sigma_\epsilon = 0.192$ (1.7 db)

$\sigma_\delta = 10°$, $\sigma_\epsilon = 0.139$ (1.2 db)

$\sigma_\delta = 15°$, $\sigma_\epsilon = 0.0695$ (0.6 db)

$\sigma_\delta = 19°$, $\sigma_\epsilon = 0$

These results assume that the main beam peak levels for independent errors directly describe antenna gain. However, as used here, true antenna gain must be defined in terms of a constant input power to the antenna. Then, in terms of this definition, the main beam peak levels will describe the antenna gain when errors are assumed that are correlated so that total power input is constant. This special type of correlation restricts the antenna gain variations so that larger errors than those quoted previously are permissible. In view of these considerations, the stated phase and amplitude error specification for the system electronics seem conservative.

Mechanizations of Self-Steerable Spacecraft Antenna Systems. There are numerous self-steering systems that have been proposed and/or built.* Some of these are mentioned in Section 4.3.5 of this study. They fall into two general classes, phase inversion types, and phase-locked loop types. A choice of circuit that would be best suited for use on a spacecraft should be made on the basis of several important factors. Among these are the signal strength at the output of each module, the percentage difference between the up-link and down-link frequencies, the complexity of the circuit, and the amount of full array

*References 36, 38, 39, 41, 43, 46, and 47.

gain needed for accurate demodulation of the information or commands on the up-link. These factors will be discussed in turn.

The signal level at the output of a module can be calculated in the usual manner. If it is assumed that a 210-foot parabolic reflector antenna with a gain of 61 db is used on the ground in conjunction with a 100 Kw transmitter operating at 2.1 GHz, and a module with a beam-width of 5° is used on the spacecraft, then, at a distance of 1 AU, the signal strength out of a single module will be -93 dbm. Since the signals transmitted to the spacecraft are generally at a relatively slow data rate, the receiving circuits of the spacecraft module can have narrow bandwidths. A narrow band receiver with a sensitivity of -93 dbm does not require extreme sophistication, and either a phase inversion or phase-locked loop type of retrodirective circuit should be acceptable. The maximum doppler shift to be expected on a mission of this type may go as high as 200 to 300 KHz. At least one of the modules must carry a master oscillator with a capability of sweeping over this range in order to acquire the signal. After lock-on, the operation would be essentially narrow band.

Array Performance with Separated Up-Link and Down-Link Frequencies. The "bandwidth" of the system, or more accurately, the percentage difference between the up-link and down-link frequencies, may be a major consideration. It is shown in Reference 30, p. 166, that a degradation in the gain of the return beam from a phase inversion type of circuit will exceed 1 db, due to scanning at a different rate than the transmitted beam, unless the following equality is satisfied.

$$\frac{\omega_e}{\omega_p} < \frac{0.263}{\sin\theta_R} \frac{\lambda_p}{L} \qquad (4\text{-}54)$$

where

ω_e = down-link angular frequency

ω_p = up-link angular frequency (pilot)

λ_p = wavelength of up-link signal

L = largest linear dimension of array

θ_R = maximum angle that beam is required to scan off the normal to the array

The assumption of a scan angle of ±2.5°, corresponding to the 3-db points of a 5-degree module, and an array size of 20 feet yields a ratio of 13 percent for ω_e/ω_p. This result is well above the ratio of 8 percent that results from the presently used DSIF frequencies of 2295 MHz and 2115 MHz. There is a possibility, however, that the up-link frequency may be changed to the 1700 MHz region. This frequency would produce a ratio of $\frac{\omega_e}{\omega_p}$ = 25 percent and the gain of the retransmitted beam would be degraded approximately 3 db.

Retrodirective circuits can be designed to work with no phase errors over any bandwidth required by the use of discrete step multiplier chains and phase-locked loops. Thus there is no degradation at all to be considered from unequal beam scanning, and arrays can be made as large as desired. A retrodirective circuit designed to eliminate phase errors will be discussed below following the treatment of other promising types.

In regards to circuit complexity, the phase inversion circuits are simpler than the phase-locked loop circuits; hence they are to be favored if they measure up to the performance requirements of the system. The problem of complexity is tied into another situation, the question of the need for full array gain on reception. In the majority of up-link transmission systems, wide bandwidth is not required since only commands are sent to the spacecraft. In this case it is likely that a single module will decode the signal with a sufficiently high degree of accuracy, and no additional circuitry will be required. Should additional wideband

modulation be required, however, such as psuedo-random noise for turn-around ranging or television pictures to a manned vehicle, it may be necessary to combine the signals from all modules to produce what has come to be termed "full array gain on receive". In that case either a more complex phase-inversion circuit or a phase-locked loop circuit must be used. The phase-locked loop circuits are usually easily adapted to provide full array gain on reception.

Promising Retrodirective Circuits. The simplest circuit that could be used for this antenna would be similar to the system of Sichelstiel et al., [37] shown in Figure 4-42. This circuit will redirect the signal back to earth with a possible loss in gain due to a beam pointing error that results from the difference between the reception and transmission frequencies as noted above. It will not add the outputs of all the modules in phase to produce full array gain on reception. Hence, the maximum signal that the receiver will obtain is that picked up by a single module.

A system that uses phase inversion by mixing and gives full array gain on reception is shown in Figure 4-43. It can be seen from the figure that a considerably greater amount of hardware is required for this system; hence, it should be used only when the additional gain on reception is absolutely necessary. Furthermore, it may be desirable to adjust other parameters in the telecommunications system so that the simpler circuit could be used. The circuit shown in Figure 4-43 would suffer the same loss in gain of the retransmitted beam due to the difference between receiving and transmitting frequencies as would the simpler circuit.

A phase-locked loop retrodirective system is shown in Figure 4-44. This system could work with very weak signals, and it is easy to obtain full array gain by the simple addition of the synchronously detected video outputs of the various elements. However, this circuit has the same problem as the phase inversion circuits: the retransmitted beam may suffer a pointing error loss due to the frequency differential.

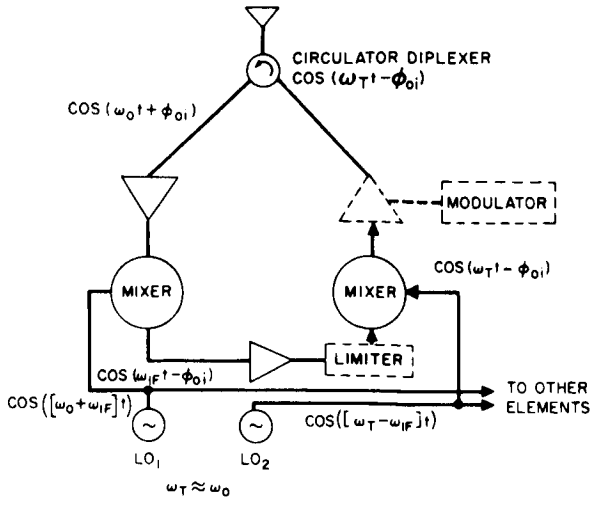


Figure 4-42. Retrodirective antenna system[37].

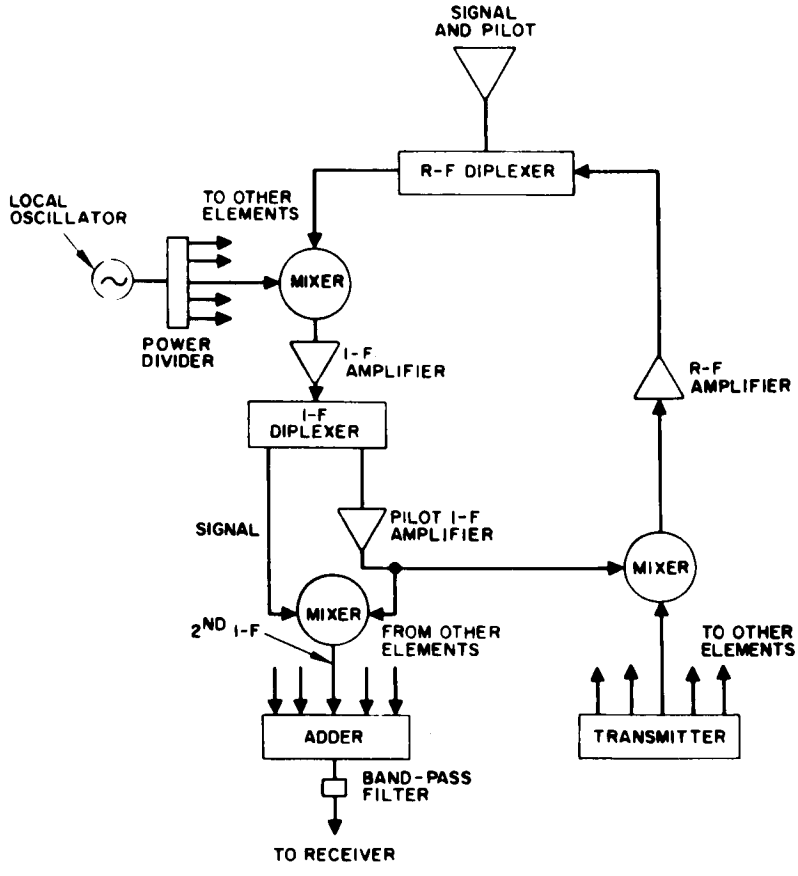


Figure 4-43. Retrodirective array using phase inversion by mixing and giving full array gain on reception.

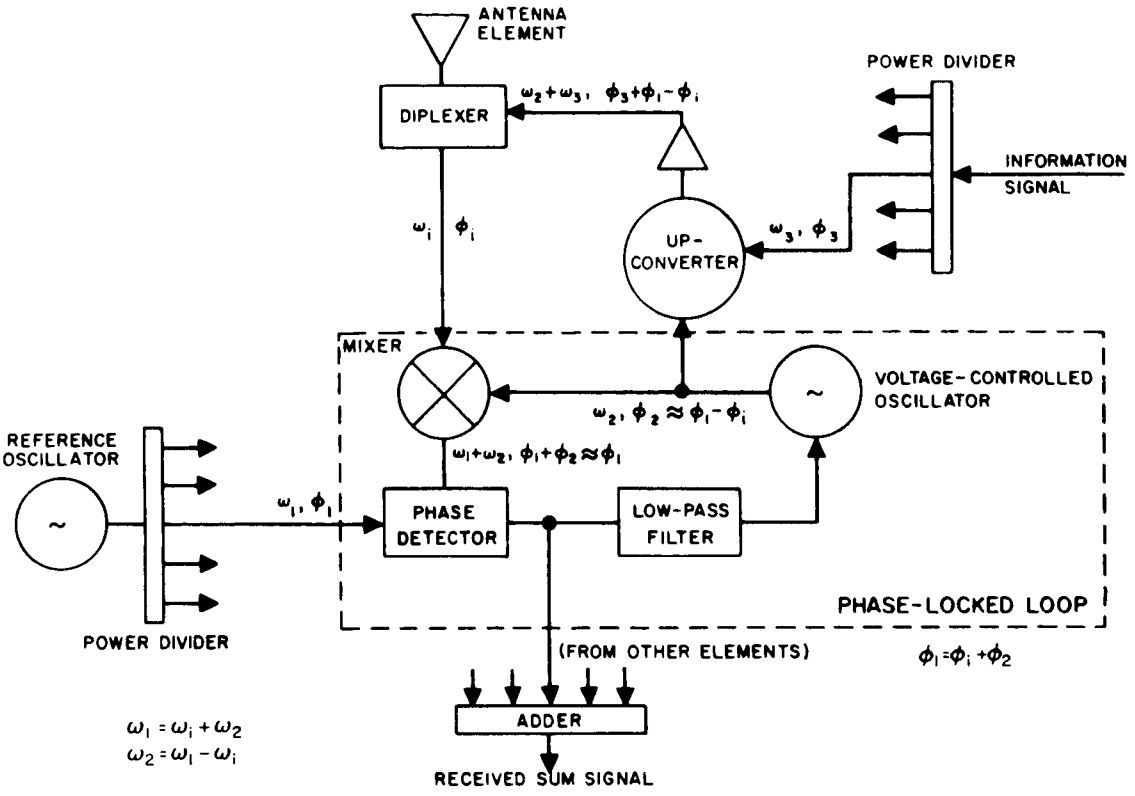


Figure 4.44. Retrodirective antenna using phase-locked loop

Arrays for Separated Up- and Down-Link Frequencies. The retrodirective circuits discussed so far are all limited in the size of the array that may be used effectively because of the different rates of scan between the receiving and transmitting beams. It would be useful to have a retrodirective circuit with the two beams that scan at the same rate regardless of the frequency difference between the transmitted and received signals. Then there would be no scanning loss to be considered, and the array could be made as large as desired consistent with other mission parameters. Further, the scan angle could be as large as desired, consistent with the limitation imposed by the beamwidth of the modules.

It is shown in the development below that the necessary condition for the achievement of equal scan rates is that the ratio of the phase difference between elements for the transmitted signal to the phase difference between elements for the received signal be equal to the ratio of the transmitting frequency to the receiving frequency.

In Figure 4-45 the geometry for determination of the scan angle of an array is shown. For the receiving frequency, the expression may be written

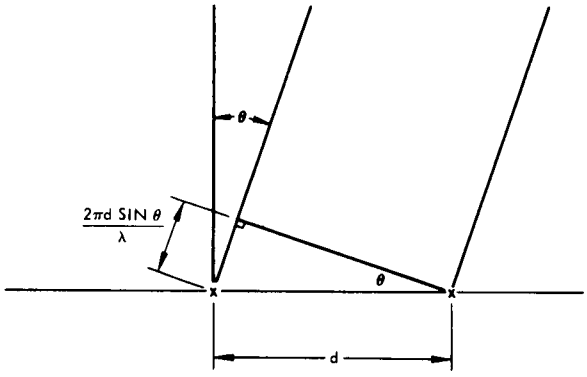


Figure 4-45. Geometry for scan angle of array

$$\phi_r = \frac{2\pi\, d \sin\theta}{\lambda_r} \qquad (4\text{-}55)$$

where

ϕ_r = phase difference between the two elements in radians

θ = scan angle off broadside

d = distance between elements

λ_r = wavelength

but

$$\lambda_r = \frac{c}{f_r}$$

where

c = velocity of light

f_r = frequency of received signal

Then Equation (4-55) becomes

$$\phi_r = \frac{2\pi\, f_r\, d \sin\theta}{c} \qquad (4\text{-}56)$$

In a similar fashion the phase difference between elements for the transmitted frequency is shown to be:

$$\phi_T = \frac{2\pi\, f_T\, d \sin\theta}{c} \qquad (4\text{-}57)$$

The ratio of ϕ_T to ϕ_r then becomes

$$\frac{\phi_T}{\phi_r} = \frac{\frac{2\pi f_T d \sin\theta}{c}}{\frac{2\pi f_r d \sin\theta}{c}} \tag{4-58}$$

Since d is constant regardless of frequency and θ should be equal at the two frequencies, these terms will cancel out along with the constants, leaving

$$\frac{\phi_T}{\phi_r} = \frac{f_T}{f_r} \tag{4-59}$$

It has been established that the transmitting and receiving beams will always point in the same direction (i.e., the scan rates are equal) when the ratio of the phase shifts between elements for the two signals is equal to the ratio of the two frequencies. A retrodirective circuit can be devised that will maintain the relationship given in Equation 4-59 and avoid the limitations that result from unequal scan rates.

The different parts of a phase-locked loop retrodirective system that meets the phase relationship determined above are shown in Figures 4-46 through 4-50. In this system the modules are not all the same size and are not all connected to the same type of retrodirective circuit. As an example, a 25-module array for use with a retrodirective system that has no scanning loss is shown in Figure 4-46. The center module is called the Reference Module and is connected to the circuit shown in Figure 4-47 which provides reference signals to the other modules so that the doppler shift can be preserved. Two pairs of rectangularly shaped modules are also shown. One pair is used to drive special circuitry which enables the system to lock-up without phase ambiguities when the received signal is off-axis in the elevation plane, and the other pair performs a similar function for the azimuth plane. These pairs are

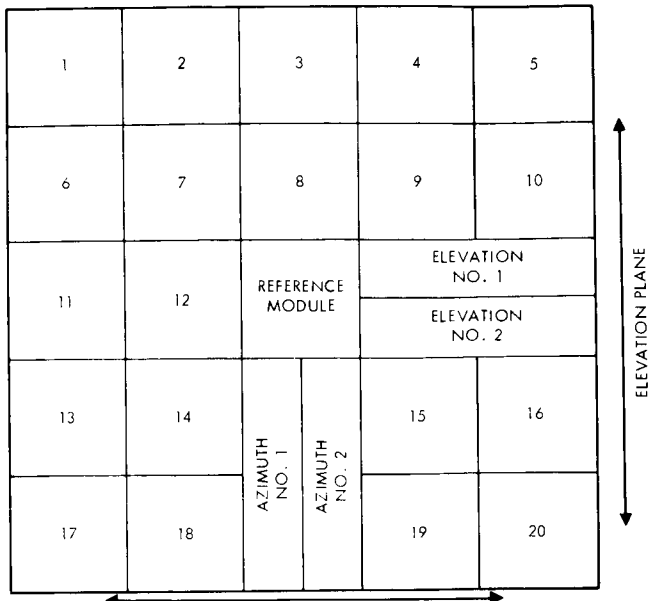


Figure 4-46. 25-module array layout

connected to the retrodirective circuit shown in Figure 4-48 and, in addition, some of their signal is used to drive the biasing circuits of Figure 4-49 and 4-50. These circuits will be described below. All the other modules are called common modules and are connected to the circuit for Figure 17 only. The operation of the various circuits will be described in turn.

The circuit which is connected to the Reference Module (Figure 4-47) operates as follows. The phase-locked loop is essentially a "divide-by-n-circuit", where n is some convenient integer, that is locked onto the received signal. The voltage-controlled oscillator (VCO) can be a transistor oscillator operating in the region from 50 to 100 MHz whose output is multiplied up to the RF operating range. This arrangement is the way voltage-controlled microwave oscillators are usually built; however, they are not always shown that way in block diagrams. In the present arrangement, part of the output of the VCO is multiplied by another integer, m (usually larger than n), for the transmitted signal. The values of m and n are selected from the numbers that can be achieved conveniently in multiplier chains and that still allow both the transmitted and received frequencies to lie in their specified frequency bands. (This technique was used in the Surveyor transponder.) It is assumed that the phase of the signal picked up by the reference module is at 0°; hence, its retransmitted signal does not have to undergo a phase inversion. The output signal is retransmitted at a frequency different from that of the received signal by a ratio of $\frac{m}{n}$, and

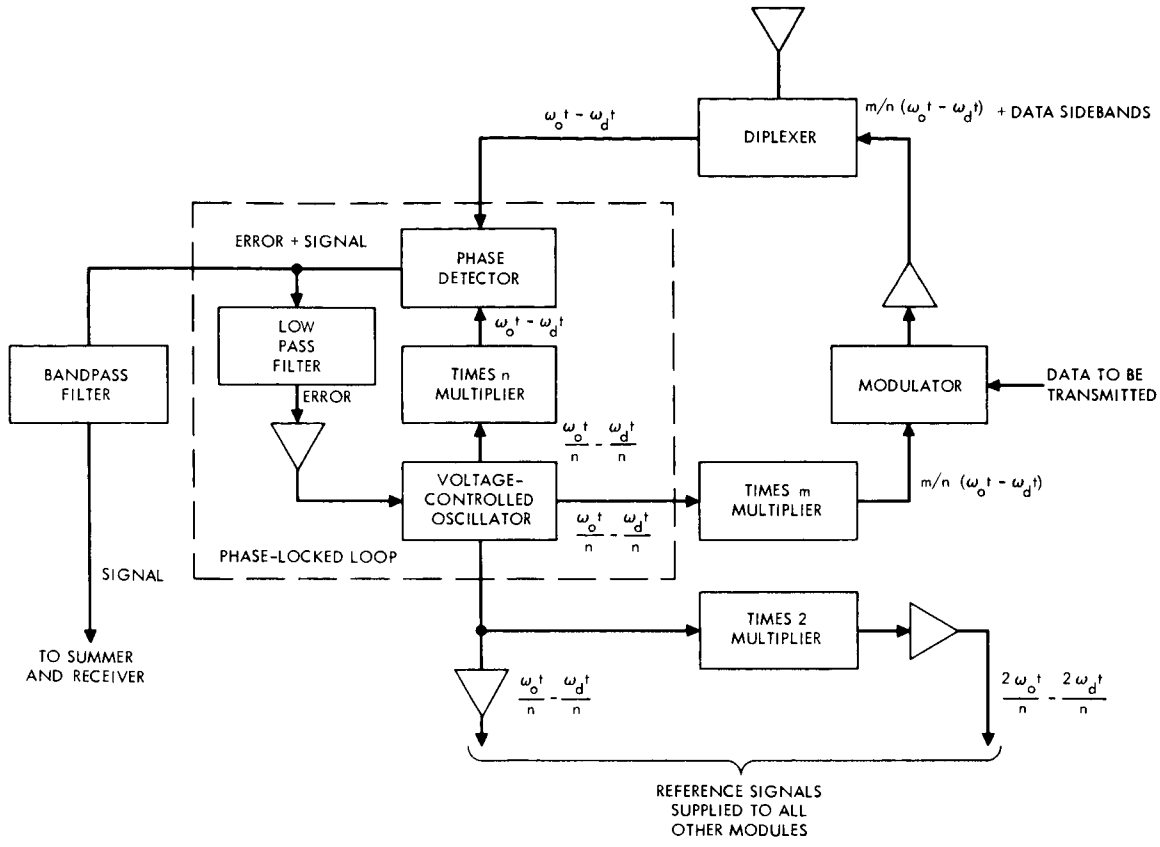


Figure 4-47. Phase-locked loop circuit for Reference Module

any component of doppler in the received signal is preserved and transmitted by the same ratio so that spacecraft velocity can be determined at the ground station.

Two signals from the VCO are amplified and sent to all the other modules. The fundamental frequency of the VCO is distributed in one line to act as the signal source for the phase-locked loops in each circuit, while the second harmonic of the VCO is distributed in another line to provide a reference signal for the mixer in each circuit. At each module (see Figure 4-48), the output of a phase detector drives a voltage-controlled phase shifter that shifts the phase of the signal from the Reference Module at the fundamental frequency so that after being

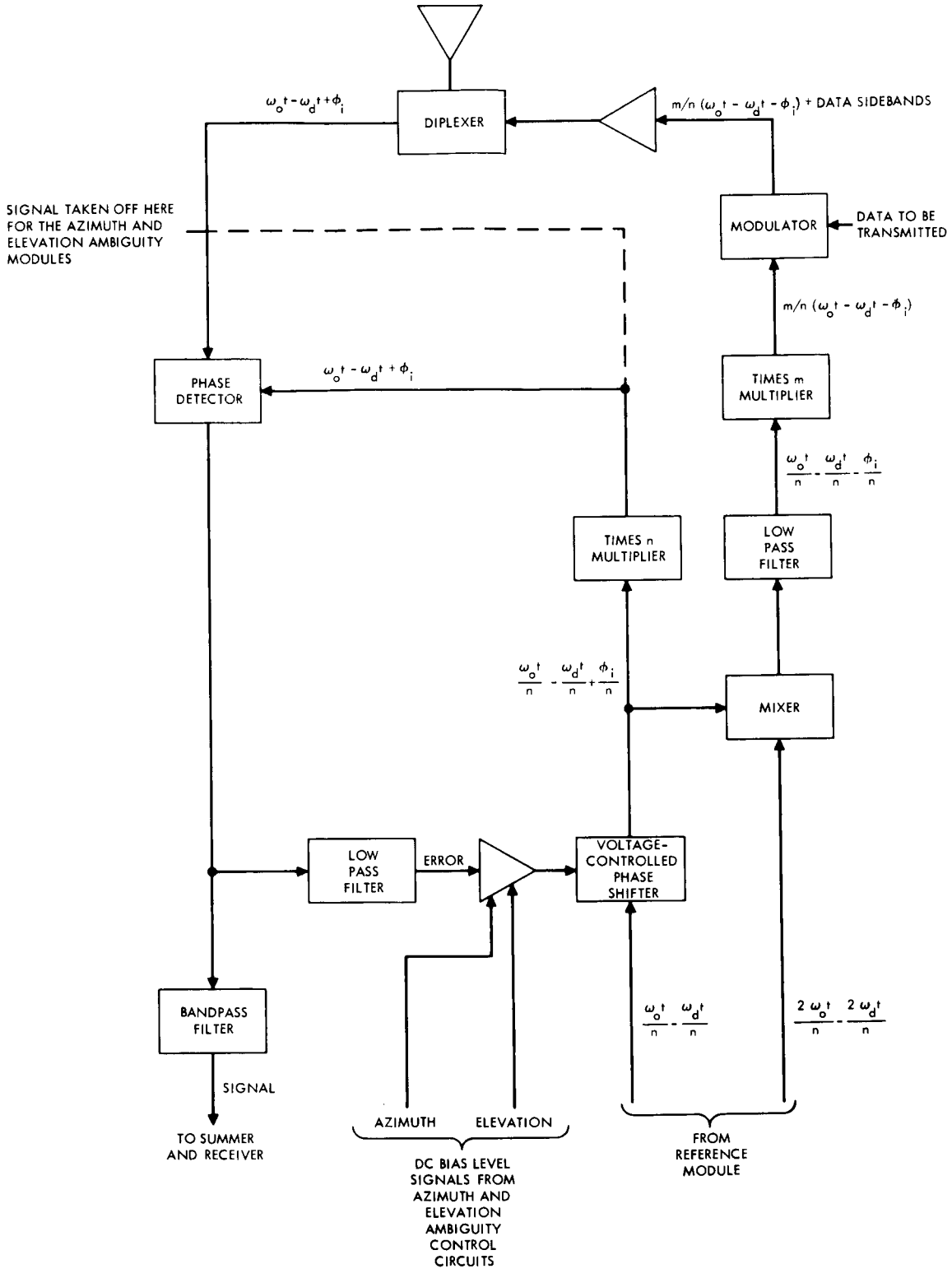


Figure 4-48. Phase-locked loop circuit for common modules

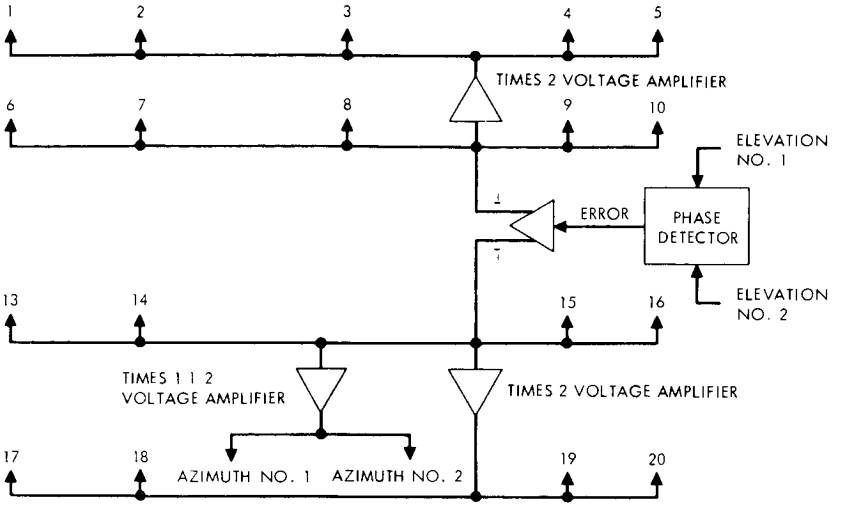


Figure 4-49. Elevation ambiguity elimination detector and bias distribution circuit.

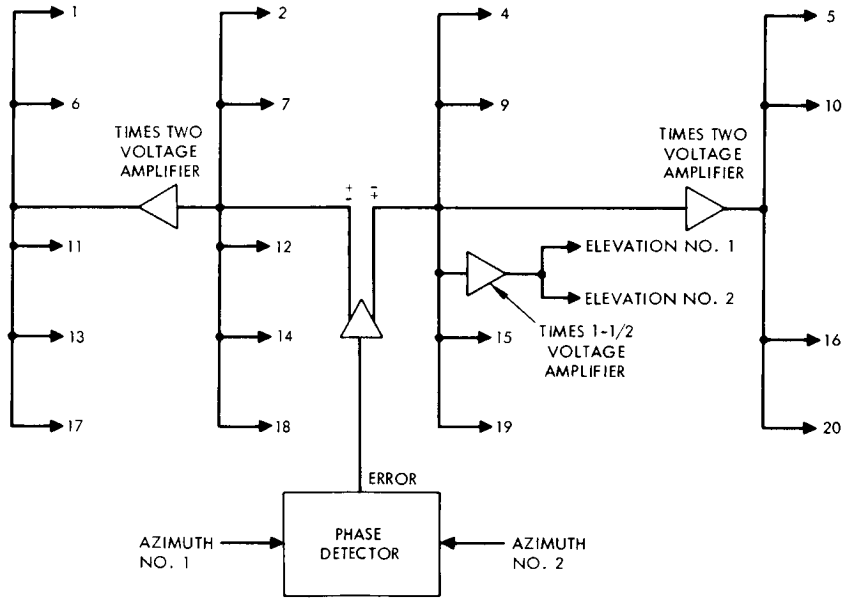


Figure 4-50. Azimuth ambiguity elimination detector and bias distribution circuit.

multiplied by n, it tracks the phase of the wave being received by that module. Thus, the signal out of the phase shifter contains a phase term unique to that module. The phase of this signal is inverted in the mixer where it is mixed with the higher frequency reference signal coming from the reference module. After the signal is filtered to obtain only the difference term and multiplied by m, it is ready to have the data modulated onto it, to be amplified, and to be reradiated by the module. It should be noted that the phase term is multiplied by $\frac{m}{n}$ so that the criteria set out above for no degradation due to scanning are met. It may also be noted that the component of received signal due to doppler is present and retransmitted also at the ratio of $\frac{m}{n}$ as was the case with the Reference Module. The retrodirective circuits considered previously in Figures 4-42 through 4-44 do not inherently preserve the doppler component of signal; to do so, they must use a reference module as described in a later discussion.

Voltage-controlled phase shifters are used in all modules, except in the Reference Module, rather than VCO's because a free running "divide-by-n" phase-locked loop could lock-up at any one of n different values of phase for its VCO. Only one of these phases will yield the proper phase for the transmitted signal after being multiplied by m. Hence, all the others are ambiguous solutions.

Use of the voltage-controlled phase shifters eliminates most of the ambiguities that can appear. Certain possibilities still exist, however, when the phase of the received signal at a common module is equal to or greater than $\pm \pi$ out-of-phase with the fundamental signal sent from the Reference Module. This problem may be overcome with a bias voltage that will shift the d-c level of the error amplifiers in the common modules by an amount sufficient to shift the phase of the reference signal so that it is within $\pm \pi$ of the received signal before the module is activated. The magnitude of the bias voltage (and its polarity) required for a signal arriving from a specific direction is determined by two phase detectors, one of which is connected to the modules labeled Elevation No. 1 and Elevation No. 2, and the other to modules Azimuth No. 1 and Azimuth No. 2 (See Figures 4-46, 4-49, and 4-50). These modules are made

long and thin so that their phase centers are sufficiently close together that the phases of their received signals differ by less than $\pm \pi$ within the scan angle anticipated (in this case $\pm 5°$).

The acquisition procedure of the array will then proceed as follows. The Reference Module is allowed to lock-up first. It then provides reference signals to all the other modules. Next the elevation modules are allowed to lock-up. They may or may not be at an ambiguous solution in regards to the azimuth plane, but this condition would not affect the phase of the output of their times-n multiplers. It would affect only their transmitted signals. The phase centers of the two modules are too close to each other and to that of the Reference Module to acquire an ambiguity in the elevation plane. The phases of the outputs of the two times-n multipliers are compared in a phase detector as shown in Figure 4-49. The output of this phase detector determines approximately the angle of arrival of the signal in the elevation plane. This output is amplified and distributed as a bias signal to all the error amplifiers of the modules whose phase centers differ from that of the Reference Module in the elevation plane. Modules on rows further from the Reference Module receive a bias voltage greater than those nearer. The bias voltages shift the d-c level of the error amplifiers, which in turn results in a phase shift out of the voltage-controlled phase shifters so that the possibility of the phase-locked loops arriving at an ambiguous position is eliminated as far as the elevation plane is concerned.

The azimuth ambiguity modules having been pre-set by the elevation bias signal are now allowed to lock-up. They cannot lock-up in an ambiguous solution because their phase centers are too close to each other and to that of the Reference Module in the azimuth plane to allow it; and the elevation bias signal precludes it in the elevation plane. The azimuth bias voltage is then generated and distributed in much the same way as is the elevation bias voltage. (See Figure 4-50.)

The next step is to turn off the elevation ambiguity modules and allow the azimuth bias voltage to preset their phase shifters. Any prior ambiguity that may have existed will thus be eliminated when they are

turned back on. The phase-locked loops of the azimuth ambiguity modules will remain locked-up if the removal and re-application of the elevation bias is not done too abruptly. This condition can be assured by the use of low pass filters on the bias lines.

When proper bias voltages are present from both the elevation and azimuth circuits, all the common modules in the array can be activated and they all should lock-on without ambiguities. This arrangement does, however, impose one limitation on the flatness of the antenna which must be kept in mind if a deployable structure is contemplated. The individual modules must not deviate from a plane by more than approximately 1/4 wavelength (about 1.3 inches at 2.3 GHz). This restriction results from the fact that the bias circuits will not be able to predict exactly the phase required at a module, and if they are allowed a tolerance of just under $\pm \frac{\pi}{2}$, then any physical misalignment that accounts for more than $\pm \frac{\pi}{2}$ phase error would produce a situation that allows an ambiguous solution. A more sophisticated circuit of this type that relates the bias voltage of each module to the received phase of adjacent modules would undoubtedly produce a more lenient flatness tolerance. It is speculated that the flatness tolerance could then be relaxed to $\pm \frac{\pi}{2}$ between adjacent modules and no overall array flatness specification would be required.

As illustrated in Figures 4-47 through 4-50 the self-steering systems are quite complex in terms of the number of components that are required for implementation. The rather complicated nature of the systems is due in a great measure to the requirement that the doppler information be preserved so that the velocity of the spacecraft can be measured and so that the beam-pointing directions coincide on both reception and retransmission. The reference element provides the signal with the doppler preserved, while the VCO's and mixers at the remaining elements provide the phase inversion necessary to transmit a signal back toward the direction of the incident signal. Each typical element has circuitry that requires the following components.

1 RF diplexer

1 phase detector

1 band pass filter (information band)

1 low pass filter (loop filter)

1 d-c error amplifier

1 voltage-controlled phase shifter

2 frequency multipliers

1 RF mixer

1 low pass filter (for RF mixer output)

1 modulator

1 RF power amplifier

The Reference Module circuitry includes the following components.

1 RF diplexer

1 phase detector

1 low pass filter (loop filter)

1 band pass filter (information band)

1 d-c error amplifier

1 voltage-controlled oscillator

3 frequency multipliers

1 modulator

1 RF power amplifier

In addition the removal of phase ambiguities as described in a previous section requires a bias distribution network.

If a relatively large number of radiating elements are to be used, then there must be available compact, lightweight components that are space qualified and that require small operating powers (high efficiency). A number of companies are developing microminiaturized and integrated components that may ultimately lead to reasonable sizes and weights of the modules. A continuing survey of the state-of-the-art in components such as these is therefore recommended as one desirable area of effort.

Since the number of components per module is quite large for an array that uses phase-locked loops for self-steering, it is important that the number of modules in an overall array be kept as small as possible. Therefore, the required value of the effective radiated power, the power available from individual RF power amplifiers, and the required

coverage region must be investigated to insure that all requirements are satisfied with the minimum number of elements. At the same time, it is also desirable to have modules with as high a gain as possible to insure that the signal-to-noise ratio at the input to the phase-locked loops is sufficiently high so that phase lock is maintained. Satisfaction of this requirement also tends to minimize the number of modules in the total array. Not all of these requirements, however, do tend to minimize the number of modules. For example, if the spacecraft orientation is not controlled sufficiently accurately, the required coverage region will be undesirably large. The module beamwidths must then be sufficiently wide to blanket the coverage region. Consequently, the module gain and size may be undesirably small. As a result, the number of modules may become excessive, or the signal-to-noise-ratio at the module may fall below the required threshold for maintaining lock.

Some typical mission parameters should be used in a realistic evaluation of the inter-relations among all these quantities.

Summary. Because of the problems associated with increasing the gain of a spacecraft antenna when the aperture size is increased, attention has been directed to self-steering arrays. Methods of effecting the design of the radiating structure are presented, together with a design example. Functional descriptions of circuitry suitable for the necessary signal processing to accomplish beam pointing with doppler corrections are given. Assessment of the state-of-the art indicates that designs can be, and in fact are being, implemented for specific spacecraft applications. Development of specific systems and components, particularly solid state amplifiers, should be continued and expanded.

4.6 REFERENCES

1. See Reference 6, Section 2.6.

2. L. E. Mathias, A. Crocker, and M. S. Wills, Laser oscillations at submillimeter wavelengths from pulsed gas discharges in compounds of hydrogen, carbon, and nitrogen, Electronic Letters 1-2, 45-46, April 1965.

3. J. B. Gunn, Instabilities of current in III-V semiconductors, IBM Journal of Res. and Dev. 8, 141-159, 1964.

4. R. I. Harrison, and J. Zucher, Hot-carrier microwave detectors, Proc. IEEE 54, 588-595, 1966.

5. S. Shapiro and A. R. Janus, RF detection by electron tunneling between superconductors, Proc. 8th Internat'l Conf. on Low Temperature Phys., pp. 321-323, Butterworth, London, 1962.

6. C. F. Krumm and G. I. Haddad, Millimeter- and submillimeter-wave quantum detectors, Proc. IEEE 54, 627-632, 1966.

7. W. G. Matthei, Recent developments in solid state microwave devices, The Microwave Journal 9, No. 3, 39-47, March 1966.

8. J. Eimbinder, Microwave IC's: hybrid or monolithic, Electronics 39, 135-142, 1966.

9. Private communication with Texas Instruments, Inc., concerning their program on Molecular Electronics for Radar Applications, Contract AF33(615)-1193, 1966.

10. P. D. Potter, W. D. Merrick, and A. C. Ludwig, Large Antenna Apertures and Arrays for Deep Space Communications, Technical Report No. 32-848, Jet Propulsion Laboratory, Pasadena, California, November 1965.

11. J. Ruze, The effect of aperture errors on the antenna radiation pattern, Supplement to Vol. IX, Del Nuovo Cimento No. 3, 1952.

12. R. S. Elliott, Mechanical and electrical tolerances for two-dimensional scanning antenna arrays, IRE Trans. AP-6, 114-120, January 1958.

13. L. A. Rondinelli, Effect of random errors on the performance of antenna arrays of many elements, 1959 Nat'l IRE Conv. Record, Pt. 1, 174-189.

14. J. I. Marcum, Table of Q Functions, Report No. RM-339, Rand Corporation, Santa Monica, California, January 1950; ASTIA Doc. No. AD-116551.

15. L. L. Bailin, Fundamental Limitations of Long Arrays, Technical Memorandum No. 330, Hughes Aircraft Company, Culver City, California, October 1953.

16. See References 37 through 41, Section 2.6.

17. See References 42 through 48, Section 2.6.

18. A. G. Smith, Extraterrestrial noise as a factor in space communications, Proc. IRE 48, 593-600, April 1960.

19. See References 7 through 9, 11 through 13, Section 2.6.

20. D. C. Hogg and R. A. Semplak, Effect of rain and water vapor on the sky noise at centimeter wavelengths, Bell System Tech. J. 40, 1331-1348, 1961.

21. R. C. Hansen and R. G. Stephenson, Communications at megamile ranges, The Microwave Journal 4, No. 12, 79-86, December 1961.

22. D. Schuster, C. T. Stelzried, and G. S. Levy, The determination of noise temperatures of large paraboloidal antennas, IRE Trans. AP-10, 286-291, May 1962.

23. N. G. Roman and B. S. Yaplee, Radio sources and the milky way at 440 MC, Proc. IRE 46, 199-205, January 1958.

24. S. Silver, ed., Microwave Antenna Theory and Design, Vol. 12, MIT Radiation Laboratory Series, McGraw Hill, Inc., New York, 1949.

25. System Capabilities and Development Schedule of the Deep Space Instrumentation Facility, 1964-1968, Revision 1, Technical Memorandum No. 33-83, p. 29, Jet Propulsion Laboratory, Pasadena, California, 24 April 1964.

26. 2.8-Minute Beamwidth, Millimeter-wave Antenna-Measurement and Evaluation, Report No. TDR-469 (5230-41), Aerospace Corporation, El Segundo, California, 26 March 1965.

27. G. Doundoulakis and S. Gethin, Far field patterns of circular paraboloidal reflectors, 1959 IRE Nat'l Conv. Record, Pt. 1, 155-173.

28. J. Stacey, Research and Experimentation of Space Applications of Millimeter Waves, Report No. TDR-169(3250-41)TN-2, Aerospace Corporation, El Segundo, California, 5 February 1964.

29. P. W. Hannan, The element gain paradox for a phased-array antenna, IEEE Trans. AP-12, 423-433, July 1964.

30. W. H. Kummer and A. T. Villeneuve, Spacecraft Antenna Systems, Interim Engineering Report (Phase I-Final Report) on Contract NAS 5-3545, Report No. P65-35, Hughes Aircraft Company, Culver City, California, January 1965.

31. C. A. Belfi and others, A satellite data transmission system, IEEE Trans. AP-12, 200-206, March 1964.

32. J. Butler and R. Lowe, Beam-forming matrix simplifies design of electronically scanned antennas, Electronic Design 9, 170-173, April 1961.

33. W. H. Kummer, Self-steering arrays for satellite applications, Section IV, Communication Satellite Systems Technology, R. B. Marsten, Academic Press, New York, 1966.

34. L. C. Van Atta, Electromagnetic Reflector, U. S. Patent No. 2,098,002, 1959.

35. R. C. Hansen, Communications satellite using arrays, Proc. IRE 49, 1066-1074 and 1340-1341 (Corrections), 1961.

36. E. M. Rutz-Philip, Spherical retrodirective array, IEEE Trans. AP-12, 187-194, 1964.

37. B. A. Sichelstiel, W. M. Waters, and T. A. Wild, Self-focusing array research model, IEEE Trans. AP-12, 150-154, 1964.

38. C. C. Cutler, R. Kompfner, and L. C. Tillotson, A self-steering array repeater, Bell System Tech. J. 42, 2013-2032, 1963

39. C. Y. Pon, Retrodirective antenna using heterodyne technique, IEEE Trans. AP-12, 176-180, 1964.

40. D. L. Margerum, Self-phased arrays, Microwave Scanning Antennas, Vol. III, Chapter 5, R. C. Hansen, ed., Academic Press, New York, 1966.

41. S. N. Andre and D. J. Leonard, An active retrodirective array for satellite communications, IEEE Trans. AP-12, 181-186, 1964.

42. E. L. Gruenberg and C. M. Johnson, Satellite communication relay system using a retrodirective space antenna, IEEE Trans. AP-12, 215-223, 1964.

43. A. F. Gangi, The active adaptive antenna array system, IEEE Trans. AP-12, 161-169, 1963.

44. F. J. Charles and W. C. Lindsey, Some analytical and experimental phase-locked-loop results for low signal-to-noise ratios, Proc. IEEE 54, 1152-1166, 1966.

45. J. E. Howard, Statistical patterns of a general array, IEEE Trans. AP-15, January 1967.

46. M. I. Skolnik and D. D. King, Self-phasing array antennas, IEEE Trans. AP-12, 142-149, 1964.

47. R. N. Ghose, Electronically adaptive antenna systems, IEEE Trans. AP-12, 161-169, 1964.

5.0 OPTICAL COMMUNICATIONS TECHNOLOGY

Optical communication links potentially capable of long-range, high data rate transmission such as will be required for deep space communications are still in the experimental stage. While the techniques and components necessary for their implementation do exist, they are at present in various phases of the research and development cycle. Because of the only-recent availability of coherent light sources (lasers) and the intense interest in the subject, developments are proceeding at a very rapid pace; so much so in fact that it is difficult to establish the current state of the art and very risky to attempt an extrapolation over a period of several years. The recent development of the CO_2 laser can be cited as an example: the first CO_2 laser, reported in April 1964, had a power output of one milliwatt; the output was increased to about the 10-watt level during the first year, and has since been increased by well over an order of magnitude. The material presented in this section is thus subject to rapid revision or complete supersedure in certain areas by new developments in techniques and materials.

5.1 OPTICAL POWER SOURCES – LASERS

Lasers present a means for the generation of coherent electromagnetic energy in the optical region. The lateral coherence of laser radiation permits focussing and collimation to the diffraction limit as from a point source, while the temporal coherence allows heterodyne (and homodyne) detection methods to be used, thus extending radio-frequency techniques into the optical region.

For transmission of data at a rate of 10^8 bits per second, the goal assumed in this study for an optical communication system, the transmitter will be effectively in continuous operation; thus the characteristics of lasers operated in the CW mode are pertinent to this application.

5.1.1 Characteristics of Materials

The active materials which are used in CW laser oscillators include gases, solids and semiconductors. While laser action has been demonstrated in materials other than those listed in Table 5-1, only the most commonly used or most promising materials are included here. Maser action has been demonstrated in the region between millimeter and infrared wavelengths but only at very low power levels as of this date.

The wavelengths available for operation comprise a set of discrete spectral lines associated with the various active materials. Generally more than one line is excited in a given laser; however, techniques exist for suppressing oscillation on all but the desired lines. Laser frequency tuning may also be accomplished, but only over a very limited range (usually less than one part in 10^4). (Nonlinear devices can be employed externally for tuning over wider ranges.)

The output powers of existing CW laser oscillators vary from a few milliwatts for single-mode operation to over 100 watts for multi-mode operation. CW powers considerably in excess of 100 watts have been realized for both CO_2 and Nd:YAG lasers. The most important factor limiting the output power is the very low energy conversion efficiency of most CW lasers. It usually ranges from 0.01 percent to 1.0 percent; however, a few exceptions are the CO_2 molecular laser (15 percent) and the GaAs injection laser (approximately 25 percent) when operated at cryogenic temperatures.

5.1.2 System Considerations

The spectral bandwidth of the laser output, which is an indication of the frequency stability, may range from a few cycles per second to several tens of gigaHertz. For multimode operation the spectral width of the output is closely related to the fluorescence linewidth of the atomic transition. Gas lasers exhibit the most monochromatic output since their fluorescence linewidths are typically one to two orders of magnitude less than for solids or semiconductors. Thermal fluctuations

Active Material	Wavelength (μ)	Output Power	Dimensions of Active Material	Comments	References
He-Ne	0.6118 0.6328 1.084 1.152	5 mW 50 mW 5 mW 20 mW	6 mm x 1.8 m	Single mode, commercially available	1
He-Ne	0.6328	900 mW	10 mm x 5.5 m	research devices	2
He-Ne	0.6328	100 mW	5 mm x 1.2 m		
Xe	3.5 9.0	0.1 mW 0.5 mW	2.6 mm x 50 cm	research device	3
Ar^+	0.4579(0.05) 0.4765(0.1) 0.4880(0.25) 0.4965(0.1) 0.5107(0.1) 0.5145(0.4)	10 W	6 mm x 60 cm	research devices, 0.1 - 0.2% efficiency	4
Ar^+	(as above)	16 W	4 mm x 2.6 m		5
Ar^+	0.4880	1 W	3 mm x 45 cm	airborne development device	6
CO_2	10.57 (0.75) 10.59 (0.25)	16 W	25 mm x 2.0 m	4.0% efficiency, single mode for each line	7
	10.59	135 W		15% efficiency	8
	10.59	300 W	4 m active length	~10% efficiency	9
Cr^{3+}	0.6943	1.0 W	2 mm x 5.08 cm	water cooled	10
$Nd^{3+}(CaWO_4)$	1.06	1 W	3 mm x 3.5 cm	methyl alcohol cooling (approximately 300°K)	11
Nd^{3+}(YAG)	1.06	200 W		0.2% efficiency	12
Nd^{3+}(YAG)	1.06	1.0 W	---	water cooled, 0.1% efficiency	13
$Nd^{3+}Cr^{3+}$(YAG)	1.06	10 W			14
$Ho^{3+}Er^{3+}Tm^{3+}Yb^{3+}$(YAG)	2.123	15 W	1-1/4 in. long	liquid Nitrogen (77°K) bath 5% efficiency	15
$Dy^{2+}(CaF_2)$	2.36	1.2 W		liquid Nitrogen (77°K) bath	16
GaAs	0.84	12 W	0.5 mm x 0.4 cm (diode dimensions)	liquid He (4°K) bath, 23% efficiency	17

Table 5-1. CW laser oscillators.

and mechanical vibrations are chiefly responsible for line broadening in single-mode operation which is required to achieve full coherence and its attendant benefits.

Production of a single-frequency output is still quite difficult because of the longitudinal mode structure of the long Fabry-Perot cavities used. Only the 10.6μ CO_2 and 3.5μ Xenon lines are narrow enough to produce a reasonable output in a Fabry-Perot resonator short enough so that only one longitudinal mode oscillates. This is due to the narrow doppler-broadened line widths of these two transitions (≈ 50 MHz for 10.6μ CO_2 and ≈120 MHz for 3.5μ Xe), but even these two transitions will require further mode selection techniques if longer, higher power tubes are considered. Because of the broad doppler line widths of the Ar and He-Ne lasers, single-frequency operation through the use of a sufficiently short Fabry-Perot resonator entails a drastic loss in output power. Techniques involving 3 mirror resonators allow the use of longer tubes at the expense of added complexity both mechanical and electronic (servo-controlled mirror positioning), but still sacrifice output power because the entire line is not used. The most promising technique developed to date is that of intracavity mode locking[18] with a subsequent coherent recombination[19] or selective output coupling[20]. This technique has been demonstrated in the laboratory, but practical power levels at a single frequency are yet to be obtained.

The beam divergence for axial (single) mode operation is characterized in the diffraction limit by λ/D, where λ is the operating wavelength and D is the diameter of the diameter of the beam. For most laser oscillators, the output beam diameter is of the order of a few millimeters, so beam divergences of the order of 0.1 milliradians can be obtained for visible light. (The semiconductor laser oscillator is somewhat unique since the active region has dimensions of the order of a few microns. The resultant beam divergence, even in the diffraction limit, is of the order of a few degrees.) For maximum output power, where nonaxial modes are excited in the oscillator, the beam divergence increases to as much as 10 milliradians. The divergence at the output

of the laser can be reduced, however, by passage through a recollimating lens system, the reduction in beam divergence being proportional to the increase in beam diameter in passing through the optical system. The price paid for smaller beam divergence is thus a larger transmitting aperture and a small loss in signal strength due to reflection and absorption in the lens system.

A laser amplifier may be employed following an oscillator to increase the available output power. By this means a high spectral radiance output can be derived from a low-power single-mode oscillator with little resultant degradation in overall efficiency due to mode selection. The relatively narrow passband of a laser amplifier requires that the operating wavelength of the amplifier and oscillator be closely matched. It is usual, therefore, to use the same laser material for both oscillator and amplifier. Even in this case, however, a temperature differential may be sufficient to put the oscillator frequency outside the passband of the amplifier. The oscillator-amplifier frequency matching problem can be alleviated somewhat by the use of laser frequency tuning techniques which allow small changes in the frequency, generally by less than one part in 10^4. Nonlinear devices may also be used to achieve frequency diversification by means of second harmonic generation; and recently a parametric laser oscillator has been reported[21], tunable ±10 percent about the neodymium wavelength. The distortion or spreading of the laser beam in passing through an amplifier depends on the optical homogeneity of the laser amplifier medium. For gas laser amplifiers this is not a problem: a diffraction limited input yields a diffraction limited output beam.

The noise generated in lasers is of three types: (1) spontaneous emission noise, (2) gain fluctuations, and (3) mode-interference noise. Except for operation near threshold, spontaneous emission noise can be neglected. Gain fluctuations due to pump power modulation can generally be reduced to an insignificant level by careful design of the pump source and associated power supplies. Mode-interference noise occurs if two or more modes are excited, as in a high power laser oscillator. The beat frequencies produced between the various modes may range

from several kilocycles per second to hundreds of megacycles per second. Schemes for phase locking the various modes by means of intracavity modulation, as mentioned above, promise to make it possible to obtain high power output with little or no mode-interference noise.

Excessive heating of laser components, as a result of the low energy conversion efficiency of laser materials, can lead to performance degradation. In a solid-state laser, for example, thermally-induced distortion, or stress birefringence, due to a nonlinear temperature distribution across the laser rod, can cause depolarization of the laser output. This depolarization is not constant but displays a radial dependence. As a result, the output from the laser has a seemingly random polarization and, where a polarized signal is required, the beam must be repolarized with consequent loss in signal power. Attention must be given therefore to methods for minimizing temperature effects, improving heat transfer, or, preferably, increasing energy conversion efficiencies to reduce input power requirements simultaneously. An improvement in conversion efficiency can result from a better match of the energy spectrum of the pump source to the excitation requirements of the material, or vice versa. For example, increased efficiency (of about 1/2 percent) has been reported [14] in a neodymium laser by the addition of chromium as a dopant to broaden the effective pump band.

5.1.3 Application to Space Communications

At present only three lasers, carbon dioxide (CO_2), "alphabet" holmium (Ho^{3+}+Er^{3+}+Tm^{3+}+Yb^{3+}:YAG) and gallium arsenide (GaAs), demonstrate the level of efficiency appropriate to spacecraft applications. While the GaAs laser is small and can utilize a low voltage supply directly for pump power, it requires refrigeration. Similarly, refrigeration is required for the optically-pumped holmium laser. The CO_2 laser, on the other hand, is somewhat larger and requires pump power at a high voltage, but because it does not need refrigeration, represents a simpler and more efficient overall system. Furthermore, it has the advantage of operating at a lower frequency in an optimal region of the spectrum, as discussed in Section 3.

Direct solar pumping of a spaceborne laser could improve overall transmitter efficiency on missions within the Mars orbit where the ambient solar flux is high. An additional complication, however, would be the need to point the solar collector at the sun while maintaining the transmitter's alignment with the earth.

The feasibility of solar pumping has already been demonstrated; continuous oscillation at 1.06 microns has been achieved for a solar-pumped Nd_2O_3-doped barium crown glass using a 60-centimeter diameter parabolic mirror[22], and for a neodymium-doped calcium tungstate laser using a cone-sphere condenser of $SrTiO_3$.[23] However, the maximum possible increase in overall efficiency is limited to the reciprocal of solar cell efficiencies; and in most cases, due to the restricted laser pump bands, the actual increase would be considerably less. Thus it is unlikely that solar pumping will have a significant influence on the development of optical communication systems.

In a ground transmitter the efficiency is not of overriding importance. Here a frequency-doubled Nd:YAG laser would certainly be a good candidate, permitting an efficient and simple noncoherent receiver on the spacecraft. Other possibilities include a phased-locked array of GaAs junctions and "alphabet" holmium, although the poorer detection efficiency at these wavelengths and the power required for refrigeration effectively cancel the improved laser efficiencies.

5.2 MODULATORS

Laser modulation is accomplished by passing the beam through an optically-transmissive medium in which one or more of the optical transmission parameters is varied by the application of a modulating field. The interaction of the laser beam and the modulating field in the presence of an optically nonlinear medium makes it possible to achieve various forms of optical modulation, including intensity, frequency, phase, and polarization.

Rather primitive devices exist at present for accomplishment of all of these forms of modulation between about 0.4 and 1.5 microns. In general, these devices are based on the use of the electro-optic effect of birefringence in crystals and can be made to provide the 100 mega-Hertz bandwidth desired for an optical space communication system. The burdens of weight and modulator driving power, however, leave much to be desired as indicated in Table 5-2. The problems associated with handling very high laser beam powers also have not yet been studied, and may present some additional difficulties. However, advances both in the synthesis of better electro-optic materials, and in the design of modulator structures should provide optical modulators suitable for space communications systems within the next few years.

For modulation in the infrared beyond 1.5 microns the best immediately available techniques involve the use of elasto-optic, or acoustic effects. Present acoustic materials and techniques can provide bandwidths of 1 to 5 megaHertz, but if 100 megaHertz bandwidths are to be achieved, it will be necessary to intensify existing research in acoustic and other means for modulation of IR lasers at wavelengths greater than 1.5 microns. While it is difficult to predict the time schedule for the availability of new electro-optic crystals, there is a good chance that materials capable of operation out to 10 microns will be found in the next few years.

Other physical processes which may be useful for optical modulation, such as controllable photon absorption in solid-state materials, and magneto-optics, are either not very promising or are in too early a state of research to accurately evaluate their potential usefulness.

Although the state of the art in optical modulators appears to be advancing at a satisfactory pace, research and development remain to be done before the technology will exist to provide optical modulators

Parameter	TM_{oo} Mode Cavity Modulator	TW Strip-Line Intensity Modulator	Zig-Zag Intensity Modulator	Strip Line Polarization Modulator (NASA Contract)	Multi-Element Intensity Modulator	Resonant SSBSC Modulator	TW Strip-Line SSB Modulator	TW Strip-Line Intensity Modulator
Development Status	Built at BTL	Built at Sylvania	Dev. at BTL	Built at Hughes Aircraft Company	Built at Hughes Aircraft Company	Built at Hughes Aircraft Company	Built at Hughes Aircraft Company	Designed at Sylvania
Material	KDP	KDP	CuCl	KDP	KDP	KDP	KDP	KDA
Crystal Dimensions	1 cm long x 2.5mm dia.	100 cm long x 2mm wide	6.7 cm long x 1mm sq	50 x 0.4 x 0.4 cm	16 each 1/4" x 1/4" x 1/2"	2 each 1-1/2" long x 1/2" dia.	80 x 0.4 x 0.4 cm	18 x 0.2 x 0.2 cm
Optical Attenuation	0.1 dB	6 dB	2.6 dB	1.5 dB	0.3 dB	1.0 dB	2.0 dB	Unknown
Useful Optical Range	0.4 - 1.5μ	0.4 - 1.5μ	0.4 - 20μ	0.4 - 1.5μ	0.4 - 1.5μ	0.4 - 1.5μ	0.4 - 1.5μ	0.4 - 1.5μ
Modulation Frequency	3 GHz	Baseband	Baseband	Baseband	Baseband	850 MHz	200 MHz	Baseband
3 dB Bandwidth	4 MHz	3 GHz	>10 GHz	30 MHz*	10 MHz*	5 MHz	100 MHz	3 GHz
Modulating Power	1.5 W	12 W	1.6 W	20 W	12 W	3 W	10 W	5 W
Modulation Index	0.13	~1.0	0.5	0.5	0.5	0.01	0.3	0.3
Extinction Ratio	Unknown	Unknown	Unknown	8:1	250:1	NA	NA	Unknown
Weight	Unknown	Unknown	Unknown	20 lbs	2 lbs	15 lbs	25 lbs	Unknown
Size	Unknown	Unknown	Unknown	144 in^3	30 in^3	850 in^3	216 in^3	Unknown

*3 dB Bandwidth limited by modulator driver and not by the electro-optic structure.

Table 5-2. Characteristics of some electro-optic modulators - June 1966.

with fully acceptable performance and burdens for space communication and tracking systems. The programs now in progress, if continued, may accomplish the desired results, but wideband optical modulation is an area which must rely on space-communications-oriented programs to provide the necessary direction and support to assure the suitability and timeliness of these results.

5.3 TRANSMITTER BEAMWIDTH – POINTING ACCURACY

The useful gain of an optical aperture (antenna) will generally be limited not by practical aperture size but rather by pointing requirements. For example, at a wavelength of 1 micron a 10-centimeter diameter aperture gives a 10 microradian beam in the diffraction limit (λ/D), equivalent to a gain of about 110 db. Since the half-angle of 5 microradians (1 arc second) represents the kind of pointing accuracy that might be anticipated for an operational system, dependent primarily on boresight and lead angle errors, optical transmitters will generally be gain-constrained. The optimum relationship between beamwidth and pointing accuracy will be determined by the distribution of pointing errors and by the dependence of communication error rate on signal-to-noise ratio.

For ground-based optical systems employing coherent detection, a somewhat more stringent limit on aperture size is set by atmospheric turbulence which, except for very carefully selected sites, will give rise to angular distortions of 10 to 100 microradians in the wavefront.

For the spacecraft transmitter, pointing errors can generally be classified into three basic categories:

- Boresight and Lead Angle Errors – These errors contribute a static or nearly constant term to pointing error and present a special problem since they cannot be enclosed within a control loop other than a DSV-earth closed loop. The penalty for excessive boresight errors is severe: the result is

likely to be loss of transmission for an extended period of time, since for ranges of 1 AU or more the round trip transit time before corrective signals can be received from the earth exceeds 15 minutes.

- Mechanical Telescope Perturbations — The steady-state response of the stabilization and tracking system may often be sufficient to suppress low frequency mechanical perturbations to a tolerable level. However, due to limited frequency response high frequency components may cause transients in the inertial telescope pointing angle in excess of desired limits. Since these errors are sensed as apparent line of sight motions by the tracker they will be reduced by the combined action of track and stabilization loops. Momentary interruption of communication or imposition of constraints on other spacecraft functions may be allowable penalties for achievement of minimum beamwidths.
- Track Errors — Angle noise in the tracker originates from such sources as the beacon tracking sensor and the inertial reference sensor of the stabilization system. Because of the low angular tracking rates required, the track loop can be given a long time constant and its noise contribution made negligible.

These categories may be expanded somewhat by examining individual pointing error causes as is done in Table 5-3.

Pointing Error Cause	Means of Correction
1. Mechanical disturbances of stabilized platform. (Bearing friction and misalignment, gimbal c.g. misalignment, spring torque from leads)	a) Attenuated by inertial stabilization b) Employ focal plane stabilization c) Separate from DSV d) Improve state-of-the-art
2. Stabilization system errors. Gyro drift (static and G-sensitive), resolver inaccuracies, accelerometer, tachometer	a) Feedback compensation b) On-gimbal star sensors c) Track loop design d) State-of-the-art improvement
3. Track errors. Sensor noise, error curve inaccuracies, resolver errors, focal plane tolerances	a) Low track loop bandwidth b) Minimize track field-of-view c) Null tracking modes
4. Mechanical alignment errors of optical axes. (Mechanical tolerance, lead angle errors)	a) Require only relative alignment b) Use closed loop where possible c) In flight alignment/ calibration
5. Atmospherics	a) Near earth relay b) Ground site selection c) Spatial averaging by distributed receivers.

Table 5-3. Pointing error causes and means of correction.

A primary limitation on pointing accuracy, and hence (except perhaps in the far IR region where aperture size and weight may be limiting) on allowable transmitter beamwidth, will be set by the accuracy to which the transmitter boresight can be maintained with reference to the tracker. Mechanisms for maintaining minimum alignment and lead-angle tolerances and methods for internal calibration must therefore be pursued.

Fine tracking can be accomplished, for example, by the movement of small elements of the system rather than movements of the entire systems. Examples are:

- Movement of the light source (or detector for the receiver) in the focal plane;
- Movement of a small transfer lens in a reimaging optical system;
- Movement of a mirror in a region of collimated light.

Methods of achieving the proper pointing vector are given in Table 5-4.

Sensing Method	Major Advantages	Major Disadvantages
Stabilized platform with gyros	Wide angle coverage; no field-of-view acquisition; rate mode for nulling tumble rates	Complex; heavy; must be updated due to drift and orbit rate; 1 to 2 year lifetime imposes questionable reliability; large
Rate-gyro stabilization	Lightweight; low power can be used for tumble-rate nulling	When used as position signal, updating required due to drift; lifetime questionable for continuous use
Ambient field sensors (gravity gradient, magnetic, etc.)	Simple, relatively passive sensors	Fields too weak and unpredictable in deep space, unless near planet
Sun sensors	Simple, reliable devices; no moving parts; flight demonstrated; lightweight; minimum power	Pointing accuracy presently limited to ±10 arcsec unless complicated horizon-scan attempted
Earth beacon sensor	Provides earth line-of-sight; applicable for coarse mode operation with gimbaled telescope; flight demonstrated	Acquisition sequence required; weight, power requirements are large compared to sun sensors; some moving parts
Star sensors	Provide accurate roll reference for coarse-fine modes; flight demonstrated	Relatively high power weight requirements; non-passive mechanically; small acquisition field-of-view; potential discrimination problems

Table 5-4. Summary of methods for achieving proper pointing vector.

Three basic types of attitude and tracking sensors are of interest.

- Sun Sensors
- Star Sensors
- Planet Sensors

Gyroscope devices are unsuitable for extended missions because of their short life (under 1000 hours typically) and unpredictable errors resulting from drift. Ambient field sensors are unsuitable because field lines are not predictable with sufficient accuracy and are too weak to be of use in interplanetary space. Approximate accuracy ranges of various sensor types are illustrated in Figure 5-1. As seen from this figure, star trackers offer high accuracy, and a great deal of development effort has been expended on such sensors. Figure 5-2 indicates the weight cost of this accuracy for a variety of star trackers.

While aperture sizes, as limited by sensor accuracy and other pointing problems, are within the state-of-the-art for established optical fabrication techniques, the weight of telescopes increases approximately as the square of the aperture diameter, and for a diffraction limited telescope with an aperture greater that 0.5 meter the weight can be expected to exceed 100 pounds even for lightweight construction employing beryllium mirrors. Hence, in the far infrared region (and specifically at 10.6 microns) a weight limitation may apply, and there will be need for a trade-off study of telescope weight, as a function of aperture and beam pointing requirements, versus transmitter and power supply weight, as a function of beamwidth. As another approach an investigation should be made of optical techniques, such as the use of a Fresnel zone plate, which exploit the monochromaticity of the signal to achieve lighter-weight systems.

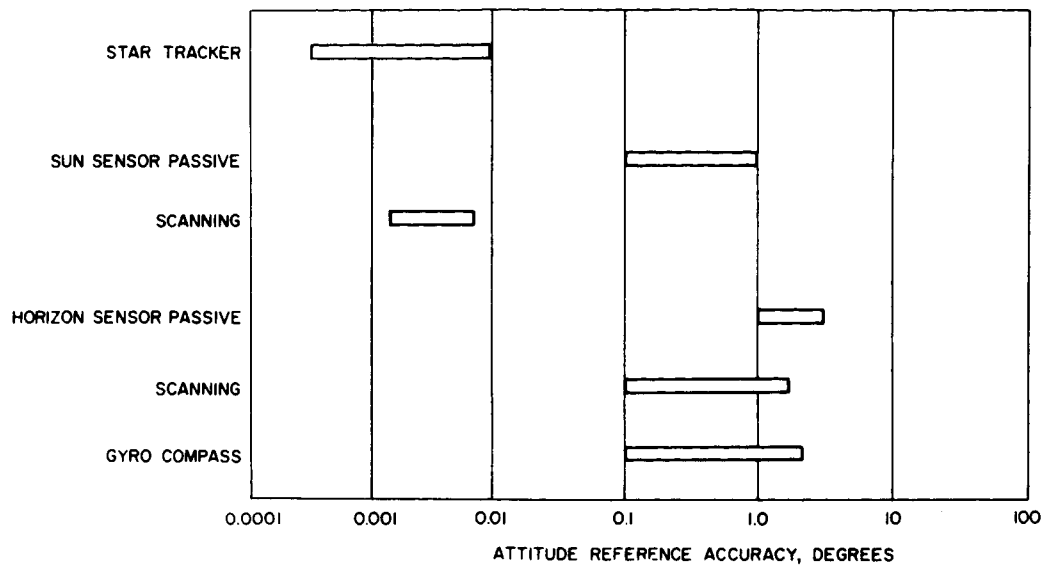


Figure 5-1. Sensor Accuracy

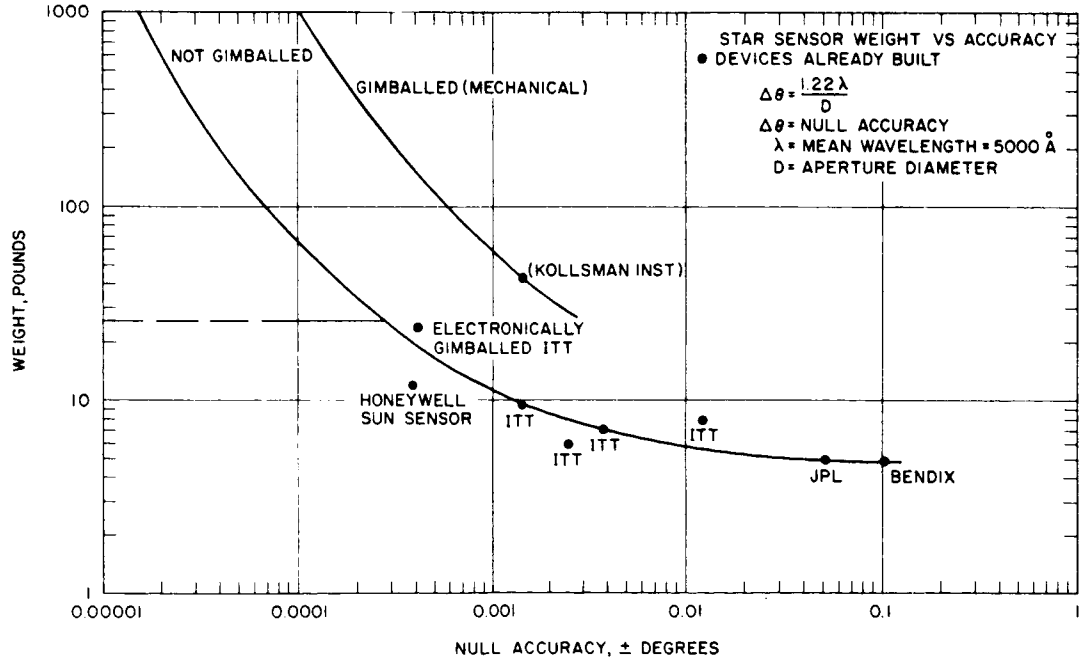


Figure 5-2. Star sensor weight versus accuracy.

5.4 RECEIVER APERTURE

5.4.1 Coherent Detection

The achievable pointing accuracy sets a restriction on the useful receiver aperture for coherent detection in a manner similar to that for gain-constrained transmission. For a coherent receiver, the pointing accuracy will determine the maximum useful aperture over which the phase of the incoming wavefront will be effectively correlated with the local-oscillator heterodyne signal. For phase correlation within a half-wavelength over the full aperture (corresponding to a maximum loss of about 3 db) the diameter must satisfy the condition $D_r \lesssim \lambda/2\delta$ so that the effective field of view corresponds to the diffraction limit of the aperture and is just $2\delta_{max} \approx \lambda/D_r$ where λ is the signal wavelength and δ is the angular pointing error. For a pointing accuracy δ = 5 microradians, the maximum effective aperture diameter at a wavelength of 10.6 microns is about 1 meter, and at 0.7 micron only 7 centimeters, dimensions which are well within the state-of-the-art for diffraction-limited optics. For a ground-based receiver, atmospheric effects will usually set a more stringent limitation by as much as an order of magnitude depending on site location (see Section 2.5 Atmospheric Distortion), where δ now represents the rms deviation in angle of the signal wavefront across the aperture.

It is evident, therefore, that for a coherent receiver, collection efficiency will be very restricted for single aperture systems. Thus, an important area for investigation will be the application of aperture-array techniques in the optical region. Self-phasing of individual apertures or, failing that, some form of postdetection correlation will be required in general for operation of optical arrays within the atmosphere.

Fine pointing may be accomplished by control of the local oscillator wavefront over a single aperture, or by proper phasing among an array of aperture elements. The latter could be most easily accomplished by means of a phase-locked loop at each element with IF signal correlation.

5.4.2 Noncoherent Detection

For noncoherent detection, where phase correlation is of no significance, aperture size is not restricted by pointing capabilities, and in fact for a ground-based receiver it is desirable to have an aperture which is large with respect to the atmospheric amplitude coherence length in order to average out scintillations. In this case "photon buckets" can be used to collect the signal over large areas so that physical restrictions and costs set the practical limit much as in the RF region of the spectrum. Requirements on overall dimensional tolerances, however, are modest for noncoherent detection and do not impose a direct limitation on effective aperture area. Rather a particular ratio of deviation in effective pathlength to aperture diameter, σ/D, corresponds to a constant field of view independent of wavelength. However, an essentially specular surface is required with angular deviations limited to the same σ/D ratio. Little work has been done on the development of large-aperture (non-imaging) light collectors. The best performance so far has been achieved for a light-weight 5.5-foot-diameter electroformed reflector[24] which had an equivalent σ/D ratio of the order of 0.5×10^{-3}. It is expected that for ground-based optics where light weight is not of prime concern this tolerance could be maintained for larger diameters and perhaps improved.

For the large apertures of interest, detectors will be either background limited (photomultipliers in the visible part of the spectrum) or detector-amplifier noise limited (solid-state detectors with little or no gain). In either case the fluctuation (shot-noise) component of the noise power (as measured at the detector output) will increase directly with aperture area while the signal increases as the square of the area (as a consequence of the square-law characteristic of photodetectors). Thus, the signal-to-noise ratio will be proportional to aperture area for a constant field of view corresponding to a particular ratio of σ/D. An increase in aperture area will always result in an increase in signal-to-noise ratio provided fabrication tolerances do not increase with diameter at a rate great enough to cause a similar increase in the effective field of view, or provided $\frac{\partial \sigma}{\sigma} < 2 \frac{\partial D}{D}$. A maximum allowable tolerance may be set, however, by a practical limitation on detector size and hence on the field of view ω; thus the deviation angle of the reflected ray must be restricted to

$$2 \frac{\sigma}{D} \lesssim \frac{\omega}{2} = \frac{d}{2f} \quad \text{or} \quad \frac{\sigma}{D} \lesssim \frac{d}{4f}$$

where f is the focal length and d is the detector diameter. Methods for construction of large "photon buckets" need study to determine the relationship between costs, aperture diameter and required dimensional tolerances.

5.5 DETECTORS

Various types of detectors have been used in the wavelength region from 0.4 to 10 microns, but many (such as thermal detectors) have too slow a response for application to wideband communications systems. For fast response, detectors based on mechanisms in which the absorbed photon energy goes into direct electronic excitation of the material are more appropriate. This category includes photoemissive detectors, such as photomultipliers, which are generally used in the visible region, and semiconductor devices, which are used both in the visible and infrared.

5.5.1 Photomultiplier Detectors

Photomultipliers are the most efficient and convenient detectors of radiation in the visible and near infrared region. Their advantage results from their fast response and the high, essentially noise-free gain achieved by secondary electron multiplication.

The process of photoemission is extremely fast so that response times of photomultipliers are limited by the time-of-flight dispersion of the electrons in the multiplier structure rather than by the nature of the detection mechanism. For structures employing focused geometries this dispersion is typically of the order of a few nanoseconds and hence bandwidths of the order of 100 megaHertz are achievable. Experimental devices employing crossed magnetic and electric field focusing and traveling wave structures have been demonstrated which are capable of even greater bandwidths.

Secondary electron multiplication raises the signal level sufficiently in respect to the thermal noise at the detector output to permit photon-limited operation. Other than shot-noise in the signal current the only noise introduced (internally) in a good photomultiplier is due to "dark current," the spontaneous emission of photoelectrons by thermal agitation. For high work-function photosurfaces as appropriate in the visible, the dark current at the cathode typically represents of the order of 10^5 electrons per second per square centimeter of sensitive surface and is thus generally negligible in relation to a signal bit rate

of 10^8 bits per second. For a low work-function surface (such as S-1) as required for the extension of the photoemissive mechanism into the near infrared, thermal emission is considerably higher; but even here, at 10^8 bits per second, the number of dark electrons per signal bit is typically less than 1 per cm^2 at room temperature and can be reduced by two orders of magnitude by cooling to -20°C.

Since the internal noise contribution is generally negligible for photomultiplier detectors, these detectors will typically be signal quantum or background noise limited. The important performance parameter is thus the quantum efficiency of the photosurface. This is given in Figure 5-3 as a function of wavelength for the better photosurface materials. The effective spectral quantum efficiency plotted comprises two components: the intrinsic quantum efficiency of the material, and the efficiency of absorption of the incident radiation within a surface layer thin enough to allow escape of the photelectroncs generated. Some enhancement can be achieved, therefore, by insertion of the radiation at grazing incidence or by causing multiple reflections within the surface in order to increase the useful path length for absorption. At the ruby wavelength of 6943Å, for example, an effective quantum efficiency of 10 percent has been achieved by this technique, as compared with the typical value of 2.5 percent. Thus, photomultiplier detectors in the visible approach within about an order of magnitude of the ideal requirement.

5.5.2 Semiconductor Detectors

The technology of semiconductor detectors has been developed primarily in response to low frequency requirements, and hence available performance characteristics are often not directly applicable to the problem of signal detection in high-data-rate systems. Response times quoted in the literature may be related either to the basic detection mechanism or to the performance of a particular detector in a particular circuit.

There are several photodiode materials with sensitivities in the visible and near infrared which have sufficiently short carrier lifetimes

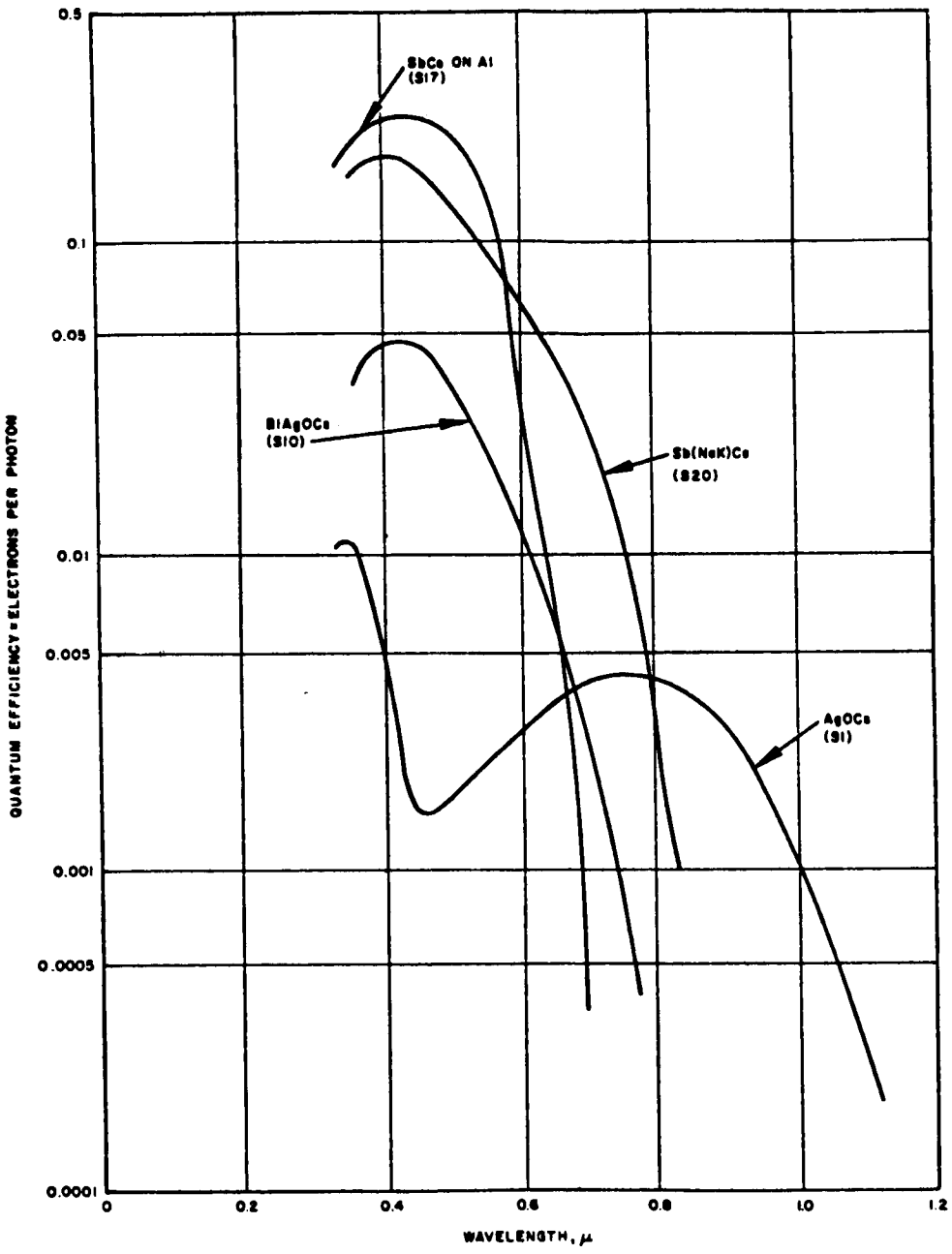


Figure 5-3. Spectral response of various photocathodes.

Laser System	Emission Line	Gallium Arsenide 300°K	Silicon 300°K	Germanium 300°K	Indium Arsenide 196°K	Indium Antimonide 77°K
Ion:Host	Microns	Structure Critical Dimension, W, and 3 db Frequency, ω, radians/second				
$Cr^{+3}:Al_2O_2$	0.6943	P-N $W_p = 2 \times 10^{-5}$ cm	P-I-N $W_I = 3 \times 10^{-4}$ cm	P-N $W_p = 2.5 \times 10^{-4}$ cm	P-N $W_p = 10^{-4}$ cm	P-N $W_p = 10^{-4}$ cm
$Sm^{+3}:CaF_2$	0.7083	$\omega = 4.7 \times 10^{10}$	$\omega = 2.9 \times 10^{10}$	$\omega = 3.1 \times 10^{11}$	$\omega = 1.1 \times 10^{11}$	$\omega = 4.6 \times 10^{11}$
$Nd^{+3}:CaF_2$	1.046			P-I-N	P-N	P-N
$Nd^{+3}:CaWO_4$	1.063			$W_I = 10^{-4}$	$W_p = 10^{-4}$	$W_p = 10^{-4}$
$Nd^{+3}:SrMoO_4$	1.064					
Ne:He	1.5			$\omega = 1.4 \times 10^{11}$	$\omega = 1.1 \times 10^{11}$	$\omega = 4.6 \times 10^{11}$
$Tm^{+3}:CaWO_4$	1.91				P-N $W_p = 10^{-4}$	P-N $W_p = 10^{-4}$
$Ho^{+3}:CaWO_4$	2.046				$\omega = 1.1 \times 10^{11}$	$\omega = 4.6 \times 10^{11}$
$U^{+3}:CaF_2$	2.49				P-N $W_p = 10^{-4}$ $\omega = 1.1 \times 10^{11}$	P-N $W_p = 10^{-4}$ $\omega = 4.6 \times 10^{11}$
λ_c microns	–	0.88	0.80	1.53	3.2	4.2
D^*_{peak}	–	2.9×10^{11}	3.4×10^{13}	10^{10}	5×10^9	1.5×10^{10}

NOTES:

1. The four upper rows are taken from Reference 24. Detectivity data as reported in various photodetector Reports, Naval Ordnance Laboratory, Corona. D* for Ge dates from 1955 and is probably not representative of best available today.
2. W_p = depth of junction behind the irradiated surface; W_I = width of the intrinsic region.
3. λ_c^p = cutoff wavelength for an absorption constant of 10^3 cm^{-1}, appropriate to wideband operation.

Table 5-5. Candidate photodiode materials for wideband optical detectors.

for use in wideband detectors. Available characteristics[25] for these materials are given in Table 5-5 and represent the best values currently achievable. There is less information on detectors for the far IR (beyond 5 microns), but available data on the most promising materials are given in Table 5-6 with spectral response shown in Figure 5-4. Response times of a few nanoseconds have recently been measured[26] for these photoconductors. Ge:Cu or Ge:Hg are currently favored for detection at 10.6 microns although multiple internal reflection techniques currently being investigated[27] for enhancement of photon absorption may extend the cutoff wavelength of Ge:Au sufficiently to permit its use at a higher temperature. The last material listed, while still in the research phase, shows great potential with good performance indicated at liquid nitrogen temperature.

The detectivity, D* (the reciprocal of the noise equivalent power for a sensitive area of 1 cm^2) is the traditional figure of merit for IR detectors. (For comparison, photomultiplier peak detectivities typically range from 10^{12} to 10^{15} $cm^{1/2}$ $cps^{1/2}$ $watt^{-1}$.) For coherent detection where internal detector noise can be made negligible, D* is not applicable; and values of quantum efficiency, the parameter of interest, have not generally been measured. The consensus is, however, that detectors can be made to give a 50 percent quantum efficiency throughout the visible and infrared spectrum. Thus, it can be anticipated that with the attention presently directed to the problem, detectors will be found that approach the ideal for coherent optical receivers.

Even for noncoherent detection, D*, since it is measured with a load resistor approximately matched to the detector impedance, bears only an indirect relationship to the effective detectivity appropriate to a wideband receiver, in which the amplifier-input resistance must be made low enough to provide the required bandwidth at the detector output. The reduction in effective detectivity in going to a 100 MHz bandwidth may be one or two orders of magnitude depending on the relative contributions of shot noise and thermal noise to the quoted D* value. Thus, presently available semiconductor detectors have a performance for wide-band, noncoherent detection several orders of magnitude below

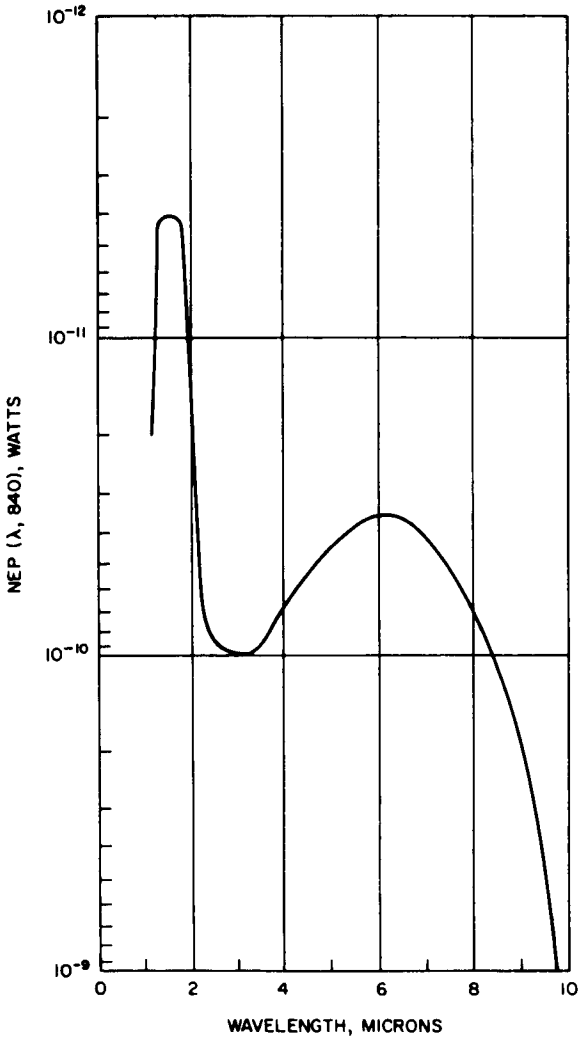
Detector Material	$\lambda^{1/2}$, μ[a]	T_{max}, °K[b]
Ge:Au	~9	70
Ge:Hg	14	40
Ge:Cd	22	25
Ge:Cu	28	18
$Hg_{1-x}Cd_xTe$	12	77

[a] Wavelength at which detectivity decreases to 1/2 its peak value.

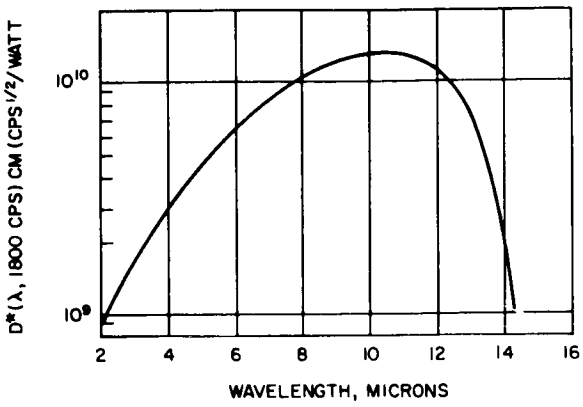
[b] Temperature at which detectivity decreases to $1/\sqrt{2}$ of its maximum value.

Table 5-6. Characteristics of semiconductor materials for 10.6 μ detector.

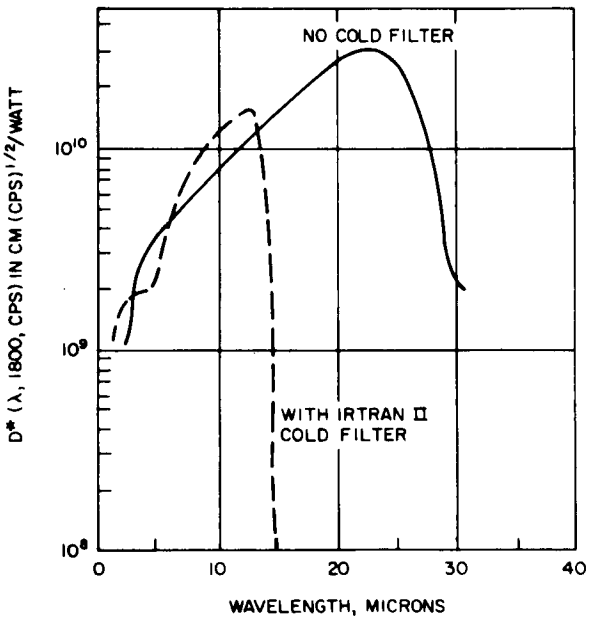
that for coherent detection, or that for photomultiplier detectors operated in either mode (assuming restricted-field, low-background conditions in the noncoherent case). Attention has been directed recently, however, to the problem of achieving signal gain in semiconductor diode detectors. Both parametric amplification and avalanche effects have been investigated[28, 29], and the latter technique has shown good promise.[30] Another technique which has demonstrated an improvement in detection efficiency is RF biasing of the detector in a microwave cavity.[31] The signal can then be coupled out capacitively, so obviating the spacecharge and spreading resistance effects associated with ohmic contacts.



a. Gold-doped germanium.



b. Mercury-doped germanium.



c. Copper-doped germanium.

Figure 5-4. Monochromatic detectivity as a function of wavelength for several extrinsic photoconductors.

5.6 REFERENCES

1) Spectra-Physics Model 125 Gas Laser, Spectra-Physics, Inc., Mountain View, California.

2) D. L. Perry, "Mirror Coating Procedures for High Power Gas Lasers," 1964 NEREM (November) paper, p. 3.5.

3) Hughes Research Laboratories Final Report, JPL Contract No. 950803 (April 1965).

4) E. I. Gordon, Private communication.

5) R. Paananen, "Progress in High Power Ionized Argon Lasers," Proceedings of the Second Conference on Laser Technology, Chicago, Illinois, 6-8 April 1965.

6) Hughes Research Laboratories Model 50 Argon Laser, W. B. Bridges, private communication.

7) C. K. N. Patel, Bell Laboratories Record, p. 311, July/August 1965.

8) P. K. Tien, Bell Laboratories, private communication.

9) J. D. Rigden and G. Moeller, Recent Developments in CO_2 Lasers, IEEE J of Quant. El., QE-2, 9, pp365-368 (1966).

10) V. Evtuhov, Hughes Research Laboratories, private communication.

11) H. R. Aldag, R. S. Horwarth, and C. B. Zarowin, The Operational Characteristics of a CW Nd: CaWO Laser in the Range of Dry Ice to Room Temperature, Appl. Optics 4, 5, 559 (1965).

12) J. E. Jackson and D. M. Yemi, High Power Laser Operation, 2nd Interim Technical Report, Contract SRCR-66-4 Linde Division, Union Carbide Corp., May 1966.

13) E. J. Woodbury, Hughes Aircraft Company, private communication.

14) R. J. Pressley and P. V. Goedestier, A High Efficiency, High Power Solid State Laser, 1965 International Electron Device Meeting, Washington, D. C.

15) L. F. Johnson, J. E. Geusic, and L. G. Van Uitert, Efficient High Power Coherent Emission from H_o^{3+} Ions in Yttrium Aluminum Garnet, Assisted by Energy Transfer, APL 8, 8, pp. 200-202 (1966).

16) J. P. Wittke, J. R. Collard, R. C. Duncan Jr., P. V. Goedertier, Z. J. Kiss, R. J. Pressley, F. Sterzer, and T. Walsh, Solid State Laser Explorations, RCA Final Report AFAL-TR-64-334, Contract AF33(615)1096, Jan. 1965.

17) KORAD Model KY-1; Korad Corp., Santa Monica, California.

18) S. E. Harris, and R. Targ, F. M. Oscillation of the He-Ne Laser, APL 5, 10, 202 (1964), S. E. Harris and O. P. McDuff, FM Laser Oscillation - Theory, APL 5, 10, 205 (1964).

19) G. A. Massey, M. K. Oshman and R. Targ, Generation of Single Frequency Light Using the FM Laser, APL 6, 1, 10 (1965).

20) S. E. Harris and B. J. McMurtry (to be published).

21) J. A. Giordmaine and R. C. Miller, Tunable Coherent Parametric Oscillation in $LiNbO_3$ at Optical Frequencies, Phys. Rev. Letters 14, 24, 973 (1965).

22) G. R. Simpson, Continuous Sun-Pumped Room Temperature Glass Laser Operation, Applied Optics, 3, 6, pp. 783-784 (1964).

23) P. H. Keck, J. J. Redmann, C. E. White, and R. E. DeKinder, A New Condenser for a Sun-Powered Continuous Laser, Applied Optics, 2, 8, pp. 827 to 831 (1963).

24) Electro-Optical Systems, Inc. Pasadena, Calif.

25) G. Lucovsky, M. E, Lasser, and R. B. Emmons, Coherent Light Detection in Solid-State Photodiodes, Proc. IEEE 51, 1, pp. 166-172 (1963).

26) W. B. Bridges, Hughes Research Laboratories, private communication.

27) P. H. Wendland and B. E. Dolratz, Multiple Reflective Laser Detector Diode, Interim Engineering Report No. 3, Contract AF33(615)-3591, Hughes Aircraft Company, 31 October 1966.

28) K. Garbrecht and W. Heinlein, Noise Performance of Photodiodes in Parametric Amplifiers, Proc. IEEE 52, 2, pp. 192-193 (1964).

29) G. Lucovsky and R. B. Emmons, Avalanche Multiplication in InAs Photodiodes, Proc. IEEE 53, 2, p. 180 (1965).

30) L. A. D'Asaro and L. K. Anderson, Avalanche Multiplication Photodiodes, Bell Laboratories Record, September 1966.

31) H. S. Somers, Jr. and E. K. Gatchell, Demodulation of low-level Broad-Board Optical Signals with Semiconductors, Proc. IEEEE 54, 11, pp. 1553-1568 (1966).

6.0 PLASMA PROPAGATION

The success of any interplanetary mission will be heavily dependent on uninterrupted communication through the terminal phase.* Various schemes have been proposed to eliminate the so-called blackout problem.[1, 2, 3, 4, 5] Although some progress has been made, no practical solution has yet been found. The purpose of this study is to evaluate briefly but quantitatively the extent of the problem for the interplanetary mission of the next decade, the time period of interest to ERC.

The final report is outlined according to the several phases of the study, as follows:

- Parametric Analysis of Blackout Problem
  - Mission Analysis
  - Entry Profile
  - Entry-Induced Plasma Sheath
  - EM Wave Propagation
  - Parametric Results and Discussion
- Refinements in the Theoretical Models
  - Trajectory Calculations
  - Flow Field Analysis and Plasma Characteristics
  - EM Wave Propagation
- Overall Assessment of the Study
- Recommendations for Future Analyses

Specific vehicles, missions, and trajectories were considered according to the anticipated requirements during the next ten years. An idealized theoretical model was formulated in order to facilitate an extensive range of parametric calculations which would serve as the basis for the general evaluation of the problem. The effect of the various

*Communication during the launch phase is degraded by signal attenuation in traversing the rocket exhaust plume. This problem is different in a number of significant respects and is not considered in this study.

simplifying assumptions involved in these results are examined and recommendations made on areas for future study.

6.1 PARAMETRIC ANALYSIS OF BLACKOUT PROBLEM

A parametric analysis of the entry-induced communications blackout problem expected in future interplanetary missions is presented. The overall investigation is described in five sections:

1. Mission analysis
2. Entry profile
3. Entry-induced plasma
4. EM wave propagation
5. Analysis of blackout data

The first four sections summarize the various assumptions and calculations made in obtaining the resultant blackout data. The results are then discussed in the final section.

6.1.1 Mission Analysis

This study area deals with the determination of likely deep space missions expected and the design parameters important to a knowledge of blackout during the atmospheric entry phase of these missions.

Unmanned missions into Mars and Venus warrant the most interest; missions involving atmospheric entry into any of the other planets appear unlikely within the time span of interest. Manned deep space missions are forecast[6] for dates well beyond the 1970's

Although return to Earth from unmanned deep space missions may occur in the future, at the present no such plans have been formulated; thus except for possible testing of interplanetary entry vehicles, there appears to be no specific need for an Earth reentry blackout analysis in connection with the particular goals of this study. Attention is therefore concentrated on Martian and Venusian entries.

Two vehicles were selected for entry analysis for each planet. The Voyager-style, blunted cone[7] was given the most study emphasis for both planets. A heavier hemispheric vehicle (similar in dimension and weight to Apollo) was also studied for Venusian entries while a scaled down non-survivable version of Voyager was included in the Martian study. Figures 6-1, 6-2, and 6-3 show these vehicles in detail.

The variables controlling blackout during entry fall into two general categories. The first may be called the entry parameters and they include:

- Initial entry angle
- Initial entry velocity
- Atmospheric model
- Vehicle ballistic coefficient (W/C_DA)
- Vehicle lift

For simplicity, only non-lifting entries are considered in the general parametric analysis. This appears to be a reasonable assumption since the present Voyager concept is a low (and possibly zero) lift vehicle.[7, 8] Variations of the other parameters were included in the study with entry angle and velocity receiving the most emphasis since they are the most flexible mission parameters.

The second category includes variables associated with the communications system. They include:

- Antenna location
- Frequency
- Polarization
- Angle of propagation through the plasma sheath

To reduce the communications parameters to two, the polariza- was assumed linear and the signal taken at normal incidence to the plasma sheath. Two antenna locations were then considered, namely, forward and side-looking. The frequency parameter was assumed to have the values 1, 10, and 94 GHz giving complete coverage of the microwave spectrum. A vhf frequency was not included because extremely severe blackout was expected in nearly all cases and therefore the data would be of little comparative value.

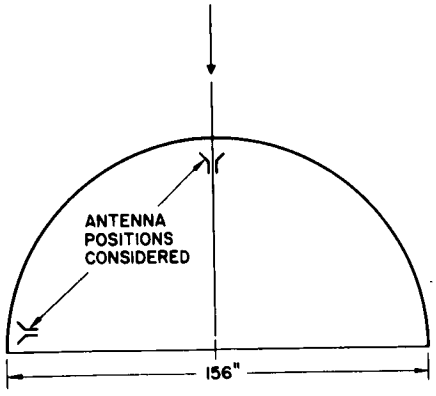


Figure 6-1. Hemispheric entry vehicle.

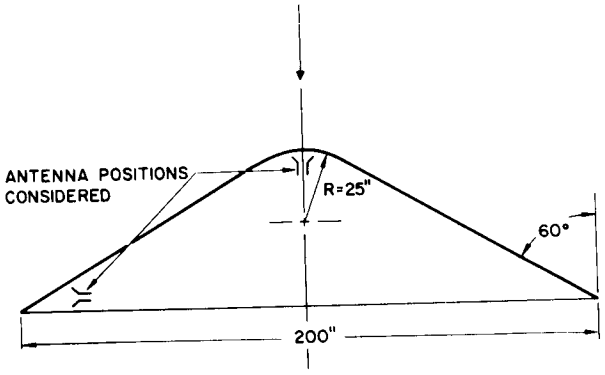


Figure 6-2. Voyager entry vehicle.

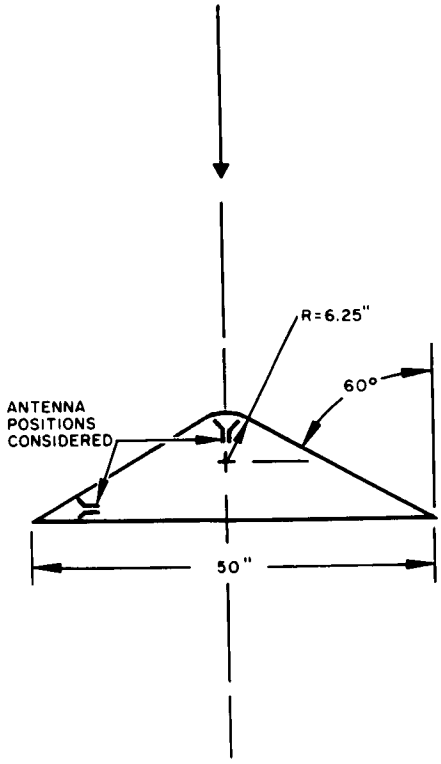


Figure 6-3. Non-survivable entry vehicle.

6.1.2 Entry Profile

The calculation of the entry trajectories was made using the HAC Planetary Glide Program which solves stepwise the equations of motion for a point mass vehicle entering the atmosphere.[9] The atmosphere density is simply given by an exponential model,

$$\rho = \rho_o e^{-h/h_o}$$

where ρ_o is the surface density, h_o is the scale height, and h is the altitude. The local gravitational acceleration is

$$g = g_o \left(\frac{R_o}{R_o+h}\right)^2$$

where g_o is the surface gravity and R_o is the radius of the planet. The equations of motion are,

$$\frac{dV}{dt} = \frac{-D}{m} - g \sin\gamma$$

$$\frac{d\gamma}{dt} = \frac{V}{r}\cos\gamma + \frac{L}{mV} - \frac{g}{V}\cos\gamma$$

where m is the mass, V is the velocity, γ is the entry angle (angle between local horizontal and velocity vector), $L = 1/2\,\rho v^2 C_L A$ is the lift, $D = 1/2\,\rho v^2 C_D A$ is the drag, with C_L and C_D being the lift and drag coefficients, respectively, and A the cross sectional area. The parameters C_D, C_L, and m and A are all assumed constant during the entry.

A number of entries were calculated for each planet from which a selection of the most satisfactory were made. This selection was based upon such practical limits as maximum deceleration, terminal velocity, and vehicle heating. The entries selected are listed in

Tables 6-1 and 6-2. It should be noted that due to the thin Martian atmosphere no entries utilizing the heavy, hemispheric vehicle ($100 < W/C_DA < 1000$) are acceptable and only low W/C_DA vehicles (Voyager = 10, non-survivable probe = 3) can be used.

In each entry approximately 12 points on the profile curve were selected for further analysis.

Vehicle	Entry Case	Entry Velocity (K ft/sec)	Entry Angle (degrees)	Comments
Hemisphere	1	15	5	$W/C_DA = 100$
Hemisphere	2	15	5	$W/C_DA = 1000$
Hemisphere	3	25	5	$W/C_DA = 100$
Hemisphere	4	25	5	$W/C_DA = 1000$
Hemisphere	5	35	5	$W/C_DA = 100$
Voyager	1	15	90	$W/C_DA = 10$
Voyager	2	25	90	$W/C_DA = 10$
Voyager	3	15	45	$W/C_DA = 10$
Voyager	4	25	45	$W/C_DA = 10$
Voyager	5	35	45	$W/C_DA = 10$
Voyager	6	15	10	$W/C_DA = 10$
Voyager	7	25	10	$W/C_DA = 10$
Voyager	8	35	10	$W/C_DA = 10$
Voyager	9	15	5	$W/C_DA = 10$
Voyager	10	25	5	$W/C_DA = 10$
Voyager	11	35	5	$W/C_DA = 10$

Table 6-1. Venus entries.

Vehicle	Entry Case	Entry Velocity (K ft/sec)	Entry Angle (degrees)	Comments
Voyager	1A	26	90	Surface pressure = 10 mb
Voyager	2A	12	90	Surface pressure = 10 mb
Voyager	3A	35	45	Surface pressure = 10 mb
Voyager	4A	26	45	Surface pressure = 10 mb
Voyager	5A	12	45	Surface pressure = 10 mb
Voyager	6A	12	5	Surface pressure = 10 mb
Voyager	1B	26	45	Surface pressure = 5 mb
Voyager	2B	20	45	Surface pressure = 5 mb
Voyager	3B	15	45	Surface pressure = 5 mb
Voyager	4B	26	20	Surface pressure = 5 mb
Voyager	5B	20	20	Surface pressure = 5 mb
Voyager	6B	15	20	Surface pressure = 5 mb
Non-survivable entry probe	1	26	90	Surface pressure = 5 mb
Non-survivable entry probe	2	15	90	Surface pressure = 5 mb
Non-survivable entry probe	3	26	45	Surface pressure = 5 mb
Non-survivable entry probe	4	15	45	Surface pressure = 5 mb

Table 6-2. Mars entries.

6.1.3 Entry Induced Plasma

The plasma and collision frequencies at points along normal lines from the vehicle surface were calculated using the HAC BLTNOS[10] program which calculates the inviscid flow-field about an axially-symmetric blunted body at zero angle of attack. The program uses the stream tube method, following lines of constant entropy from points just back of the shock wave. Real gas effects are considered with all the important species included. The gas is assumed to be in chemical equilibrium at all points as a simplifying approximation.

For the vehicles studied the desired quantities were analyzed along normal lines spaced at 5^{o}, 20^{o}, 50^{o}, and 80^{o} measured from the apex. However, only the 5^{o} and 80^{o} data was used (corresponding to the forward and side looking antenna locations).

The electron concentrations at each point studied on a given normal line were then computed using an accurate composition program which accounts for contributions from many ionization reactions as well as diminutions by two electron capture reactions. The electron collision frequencies were computed with the use of several collision cross-sections, to account for the several possible types of collision.

The results of these calculations were then used in the following analysis.

6.1.4 EM Wave Propagation

For the type of vehicles studied the plasma sheath created in front of the forward-looking antenna is very nearly homogeneous while the sheath in front of the side-looking antenna is quite inhomogeneous. In the first case the blackout can be computed using the boundary value solution of a normally incident EM wave upon a homogeneous plasma slab,

$$\text{Attenuation (db)} = 20 \log \left| \frac{4K^{1/2} \exp(ik_o Z)}{(K^{1/2}+1)^2 \exp(ik_o K^{1/2} Z) - (K^{1/2}-1)\exp(-ik_o K^{1/2} Z)} \right|$$

where Z is the plasma (shock layer) thickness and k_o is the free space wave number. K is the complex dielectric constant of the plasma,

$$K = 1 - \frac{(\omega_{p/\omega})^2}{1 - i\nu_{c/\omega}} \quad ,$$

where ω is the propagation frequency, ν_c is the collision frequency, and the plasma frequency, ω_p, is given by $\omega_p = 5.8 \times 10^4 n_e^{1/2}$ for an electron density n_e.

A more complicated calculation must be made to find the attenuation for the inhomogeneous plasma. Here $n_e = n_e(z)$, $\nu_c = \nu_c(z)$, and therefore, $K = K(z)$ (where z is the distance normal to the surface within the plasma). The solution of this problem for the arbitrary plasma distribution encountered in reentry requires a machine solution of the following equation,

$$\frac{d^2E}{dz^2} + k_o^2 K(z)\, E = 0.$$

This equation is valid only for normal incidence upon the plasma sheath (the case studied).

6.1.5 Analysis of Blackout Data

Signal attenuation calculations are presented and discussed in this section as a function of the various parameters. A typical plot of attenuation versus altitude is shown in Figure 6-4 for the Venus-Voyager entry 1 (the entry profile shown in Figure 6-5), for the forward antenna at 10 GHz. An important general feature of the problem is thus demonstrated, namely, that the attenuation is characteristically nearly always abrupt. The transition from essentially lossless transmission to a very high level of attenuation occurs over a relatively short interval in altitude and time. In the illustrative example it is interesting to note that the peak value in excess of 100 db although it occurs during only 20 percent of the altitude range or about 25 percent of the time

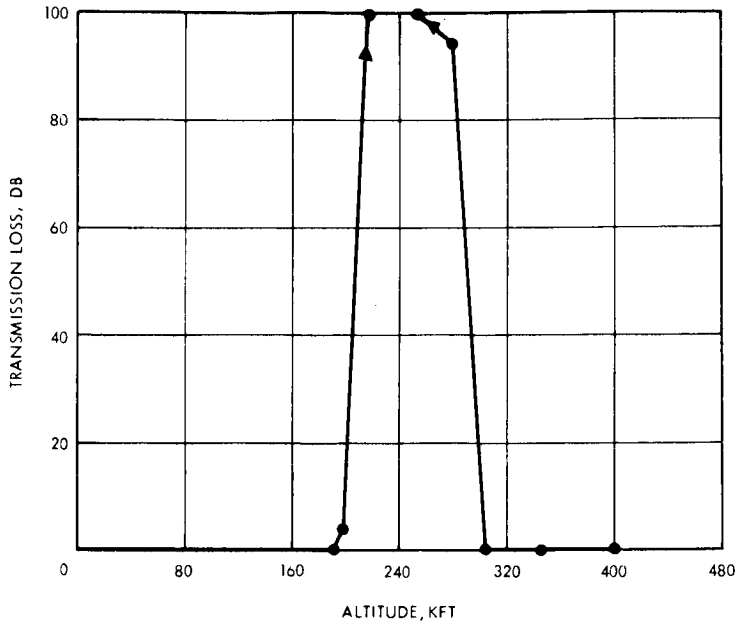


Figure 6-4. Venus-Voyager, Entry 1, forward antenna at 10 GHz.

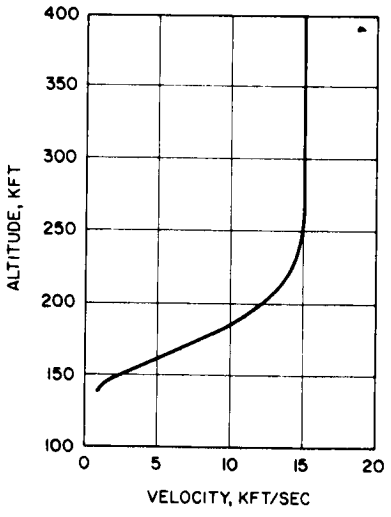


Figure 6-5. Venus Entry 1.
$\gamma_E = 90^\circ$
$V_E = 15$ K ft/sec
$L/D = 0$

(see Table 6-3). Nearly every case investigated showed a similar, abrupt behavior.

The abruptness of the blackout phenomenon leads to the conclusion that the establishment of a precise level of attenuation above which the system is considered blacked-out is not of great significance. A conservative criterion has been adopted, therefore, that a communications link will be considered to be blacked out whenever the attenuation exceeds 3 db. In Figure 6-4, for example, the error introduced in the blackout altitude band by assuming a 3 db cutoff when the system could actually tolerate 30 db attenuation is about 15 percent. Such an error is comparable with (if not substantially smaller than) the errors arising from the use of the approximate model for the parametric analysis, as will be noted subsequently.

It is clear that the precise level of attenuation at any instant is of less interest than the overall time the system is blacked out and that any alleviation of the problem must manifest itself by a reduction in the total blackout time. A reduction in the degree of attenuation at any instant means little if it does not significantly alter the time in blackout. A knowledge of the altitude bounds of blackout is also of interest, particularly when maneuvers are to be made during entry or when real time measurements are to be made. (Suggested plans for a non-survivable entry probe to measure the Martian atmosphere in 1969-1971 can be expected to encounter such difficulties.) Thus the parametric calculations of attenuation are presented in terms of the blackout time relative to overall entry time (Tables 6-3 and 6-4), and in terms of altitude bands, via bar graphs (Figures 6-6 through 6-15).

Blackout as a Function of Time

Table 6-3 lists the maximum total blackout times for the 16 entry cases for Venus while Table 6-4 presents the 16 cases for Mars. The entry times shown in both tables are the total entry times from the start of the sensible atmosphere (500,000 feet for Venus, 700,000 feet for Mars) to a limiting terminal velocity (1,000 fps) or altitude (5,000 feet). The only altitude limited terminations occur in the Mars

Vehicle	Entry Case	Entry Time (sec.)	Blackout Time					
			Forward Antenna			Side Antenna		
			1 GHz (sec.)	10 GHz (sec.)	94 GHz (sec.)	1 GHz (sec.)	10 GHz (sec.)	94 GHz (sec.)
Hemispheric	1	235	100	89	0	0	0	0
Hemispheric	2	241	106	96	44	44	0	0
Hemispheric	3	225	120	114	62	120	88	0
Hemispheric	4	249	160	132	126	142	128	110
Hemispheric	5	290	206	198	180	206	198	140
Voyager	1	35	12	8	0	20	14	10
Voyager	2	18	12	10	6	16	14	9
Voyager	3	48	15	10	0	28	17	10
Voyager	4	33	16	14	6	20	19	13
Voyager	5	25	17	17	12	17	17	15
Voyager	6	146	45	28	0	76	54	28
Voyager	7	118	60	48	17	76	60	43
Voyager	8	104	58	58	40	61	60	55
Voyager	9	219	67	20	0	117	74	20
Voyager	10	201	108	80	0	144	120	66
Voyager	11	188	132	132	90	135	138	120

Table 6-3. Maximum total time (seconds) in blackout-Venus entries.

Vehicle	Entry Case	Entry Time	Blackout Time					
			Forward Antenna			Side Antenna		
			1 GHz (sec.)	10 GHz (sec.)	94 GHz (sec.)	1 GHz (sec.)	10 GHz (sec.)	94 GHz (sec.)
Voyager	1A	33	23	22	10	30	26	18
Voyager	2A	67	8	0	0	20	14	0
Voyager	3A	41	30	30	22	34	34	30
Voyager	4A	56	32	27	12	40	32	25
Voyager	5A	106	0	0	0	33	30	0
Voyager	6A	676	0	0	0	0	0	0
Voyager	1B	48	34	30	20	38	35	25
Voyager	2B	61	35	25	15	40	37	25
Voyager	3B	80	32	20	0	46	35	20
Voyager	4B	169	73	65	25	85	78	53
Voyager	5B	190	70	47	0	80	73	35
Voyager	6B	215	60	40	0	90	65	20
Non-survivable entry probe	1	49	20	8	6	20	21	14
Non-survivable entry probe	2	68	20	16	0	20	20	0
Non-survivable entry probe	3	67	23	27	0	28	28	20
Non-survivable entry probe	4	87	25	0	0	25	25	0

Table 6-4. Maximum total time (seconds) in blackout-Mars entries.

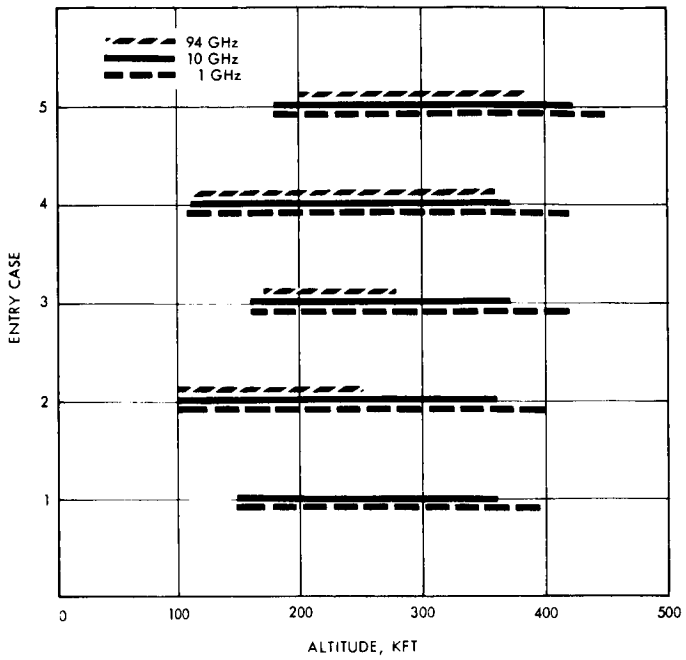


Figure 6-6. Altitude bands of blackout, Venus-hemispheric vehicle, forward antenna location.

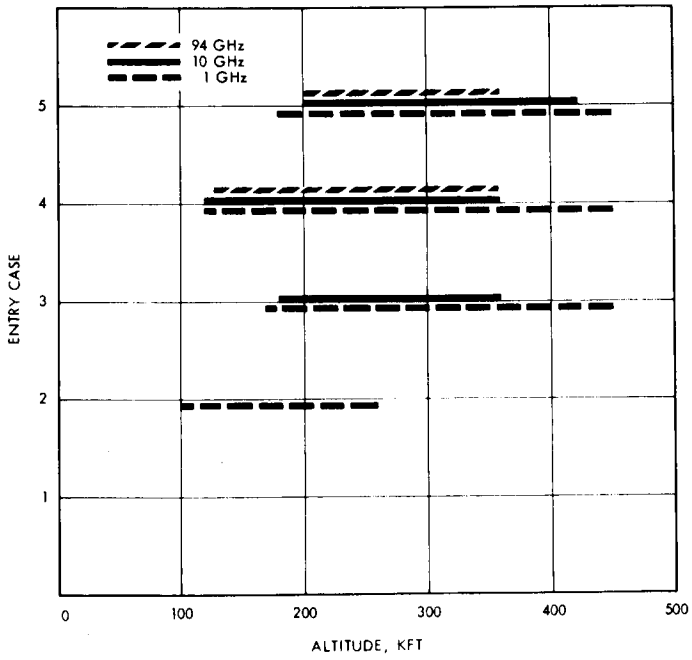


Figure 6-7. Altitude bands of blackout, Venus-hemispheric vehicle, side antenna location.

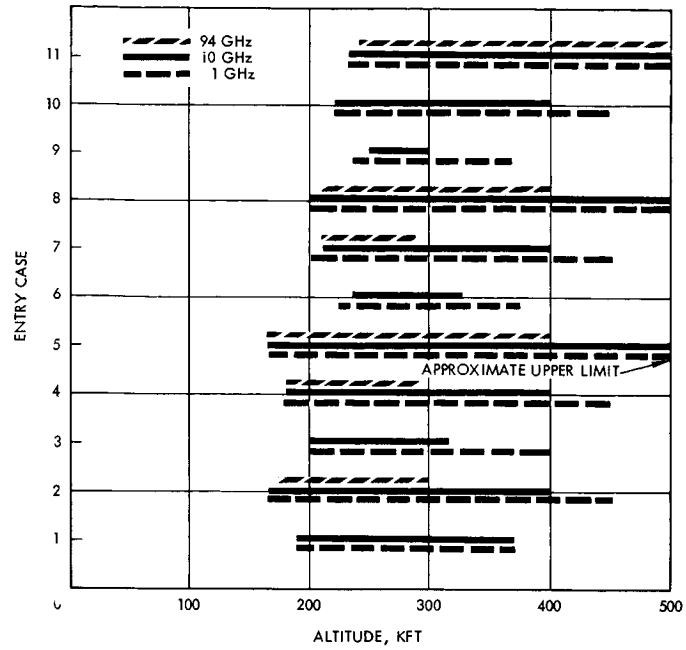


Figure 6-8. Altitude bands of blackout, Venus-Voyager, forward antenna location.

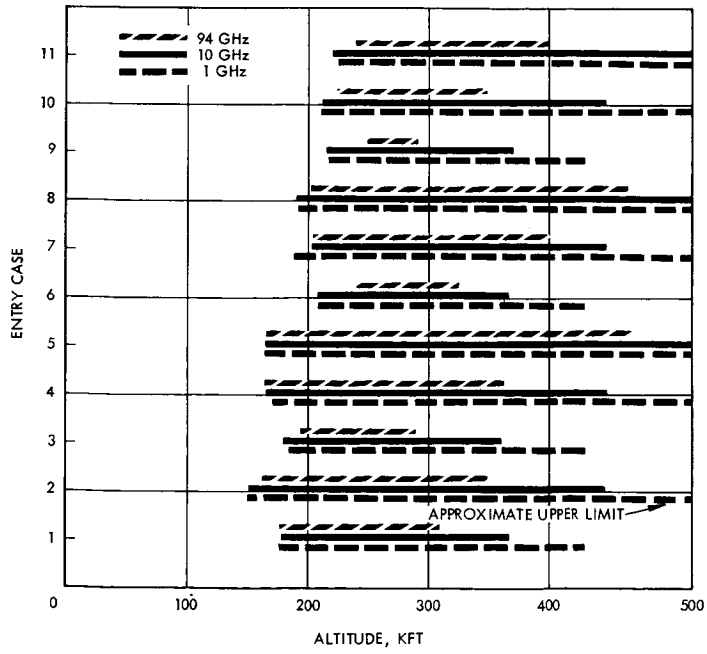


Figure 6-9. Altitude bands of blackout, Venus-Voyager, side antenna location.

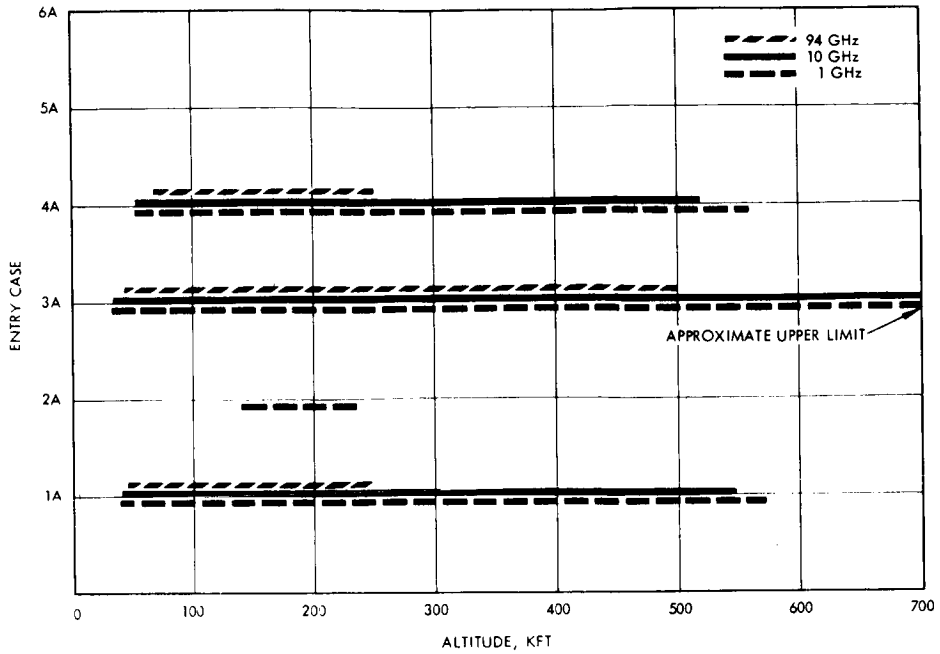


Figure 6-10. Altitude bands of blackout, Mars-Voyager, 10 mb atmosphere, forward antenna location.

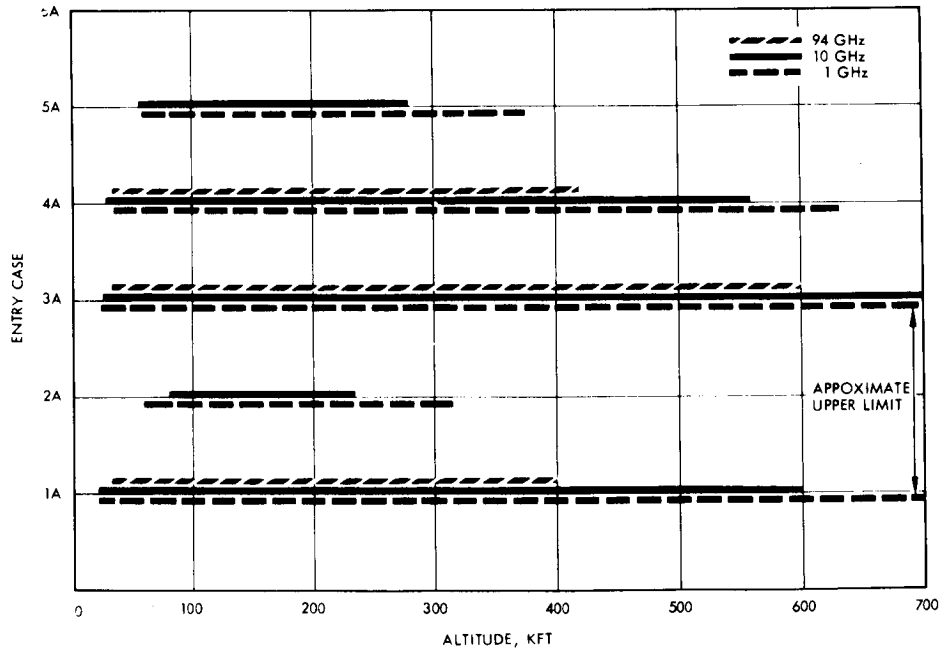


Figure 6-11. Altitude bands of blackout, Mars-Voyager, 10 mb atmosphere, side antenna location.

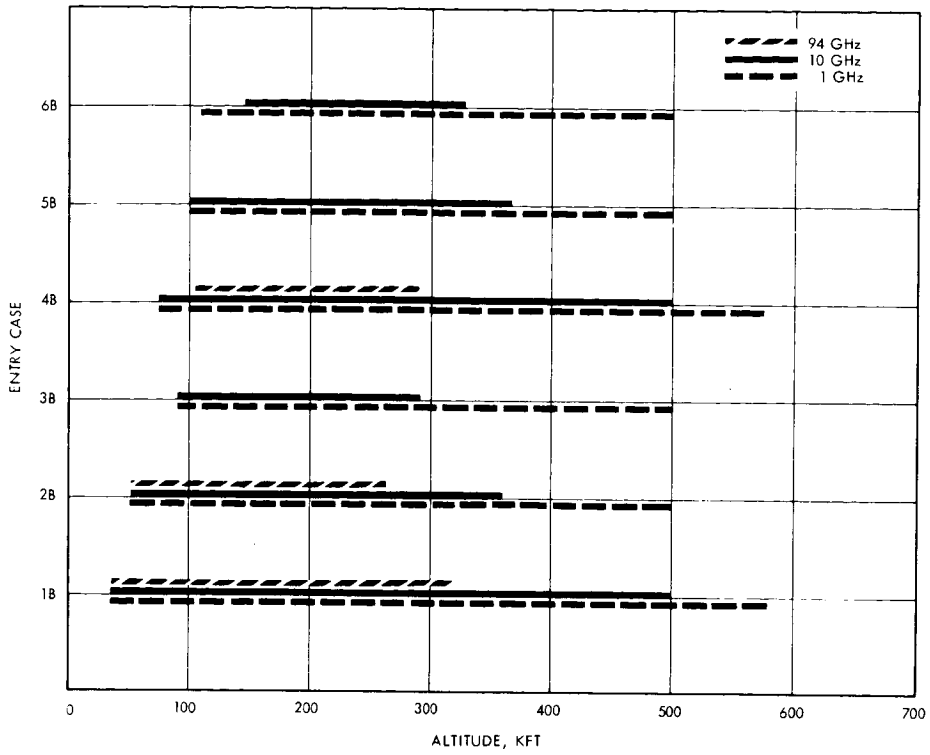


Figure 6-12. Altitude bands of blackout, Mars-Voyager, 5 mb atmosphere, forward antenna location.

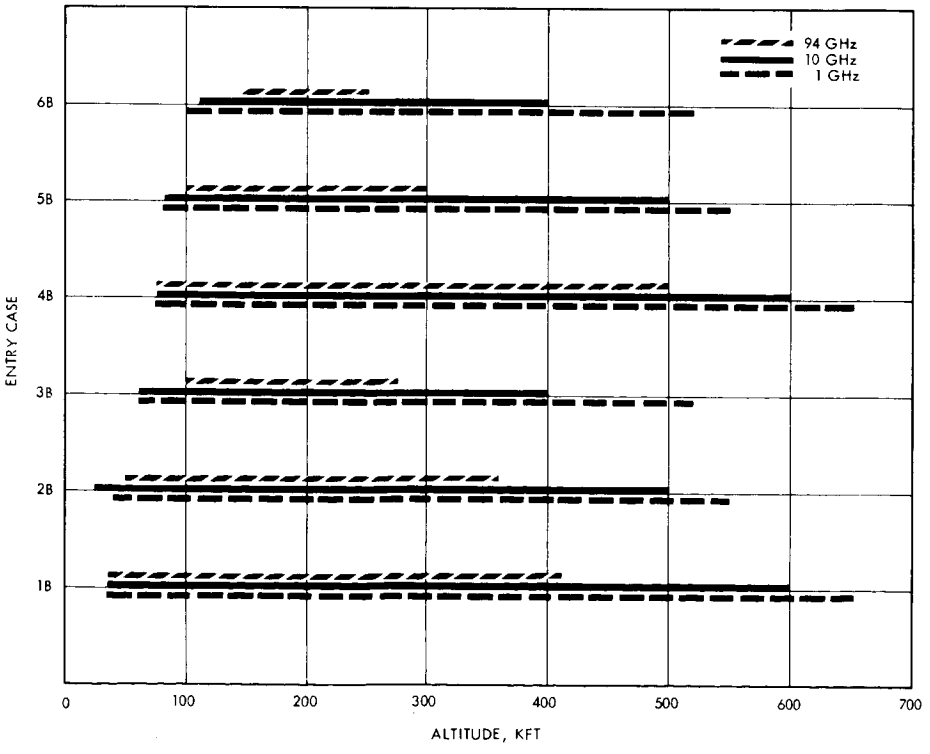


Figure 6-13. Altitude bands of blackout, Mars-Voyager, 5 mb atmosphere, side antenna location.

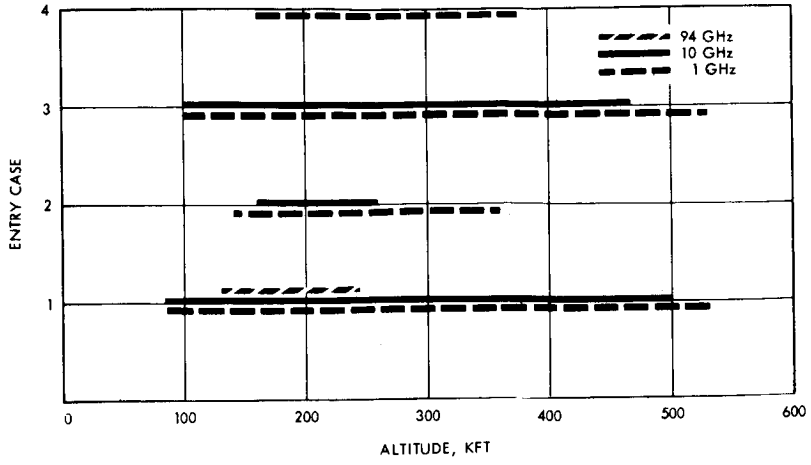


Figure 6-14. Altitude bands of blackout, Mars-non-survivable probe, forward antenna location.

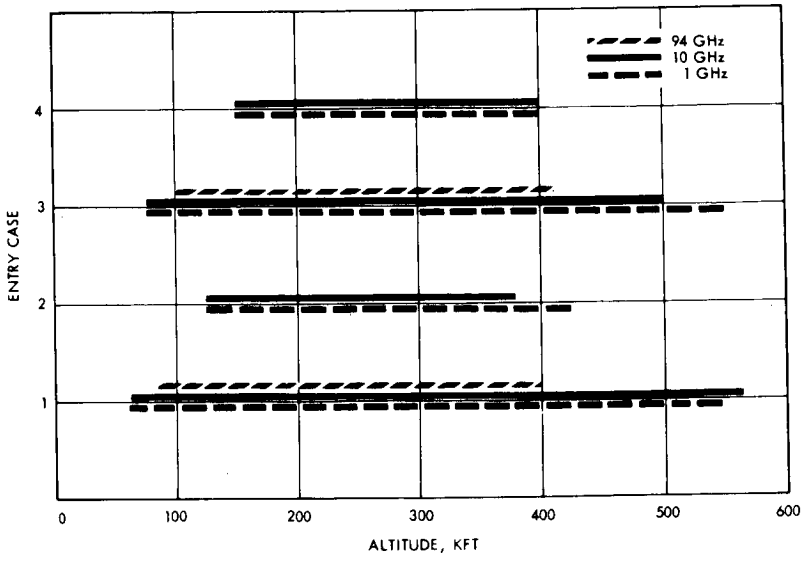


Figure 6-15. Altitude bands of blackout, Mars-non-survivable probe, side antenna location.

Voyager entries: 1A, 2A, 3A, 4A, 5A, 1B, 2B, 3B. All the remaining cases are velocity limited.

Frequency. In general, the blackout times vary inversely with the frequency, as expected. This is not always the rule, however, as seen in cases of Venus Voyager entry 11 and the non-survivable probe entries 1 and 3. The plasma attentuation factor per unit of electrical length is inversely proportional to the square of the frequency, but the electrical path length through the plasma is directly proportional to frequency. Under certain conditions this can lead to a reversal of the general trend, as shown. Furthermore, although a useful improvement often results from an increase in frequency, this is not always the case, and the effect must be determined for the specific set of mission parameters under consideration. No general rule of thumb can be applied.

Antenna Location. Blackout is a definite function of the antenna location, however, the dependency is not always clear a priori. The reason for this fact is the offsetting effects of electron density and thickness of the layer. In the case of the hemispheric vehicle, for example, a large improvement in transmission is seen for the side antenna over the forward antenna at 15,000 fps entry velocity (cases 1 and 2), less improvement at 25,000 fps (cases 3 and 4) and nearly none at 35,000 fps (case 5). Where the blackout time is decreased by using the side looking antenna in the hemispheric vehicle the reverse is clearly true for the Voyager type vehicle entering Venus. The forward antenna location is also more desirable for the Voyager entry into Mars, but to a lesser extent. It is apparent that, as in the case of frequency selection, optimum antenna location can only be determined for each specific mission and no general rule can be derived.

Entry Parameters. General rules can be formulated with reference to the entry parameters. Specifically, the relative blackout time, i.e., the ratio of blackout time to entry time, increases with increasing entry velocity and angle. Such is also the case with increasing weight (variations with $W/C_D A$ correspond to a change in weight since C_D and A are fixed by the assumed vehicle geometry). It is apparent

that the entry parameters can affect blackout substantially and must be considered in any general evaluation of the problem. Although no attempt is made to treat the effect of the atmospheric model parametrically, this factor can be of importance, as shown by a comparison of cases 4A and 1B (identical entry velocity and angle).

Blackout as a Function of Altitude

The bar graphs (Figures 6-6 through 6-15) show the maximum altitude blackout bands for the 3 frequencies in each case studied. Although band widths vary in the same qualitative manner as the blackout times discussed above, several interesting points are noteworthy:

Altitude Limits. The low altitude limit where the vehicle comes out of blackout is very insensitive to frequency changes in nearly all cases, and, in lesser degree, to antenna location. This is due to the rapid decrease in kinetic energy of the vehicle which results in a large reduction in the attenuating properties of the plasma in a relatively short altitude span. The change in plasma character is so rapid and complete that all the frequencies appear to transmit at the same altitude. This is not true at the high altitude limit where the velocity is relatively constant and the collision frequency is small (the latter increases almost exponentially with decreasing altitudes). The plasma effect is almost regular in predictability; for frequencies below plasma frequency there is good transmission, above there is evanescence, and therefore attenuation is primarily a function of the increasing gas density.

Entry Angle. Decreasing the entry angle tends to raise the low altitude limit. This is due to the fact that as the angle decreases the major portion of entry deceleration occurs at higher altitudes. The kinetic energy of the vehicle is expended at these higher altitudes and ionization in the sheath diminishes earlier. The entry angle has little effect upon the high altitude limit since in most cases blackout occurs before the start of deceleration so that the controlling entry parameter is the initial entry velocity.

Entry Velocity. Both the upper and lower altitude limits vary directly with entry velocity in the expected manner, as shown by the data.

The data presented here are intended to indicate, over a very wide range of expected missions, the general trends characteristic of atmospheric entry into the near planets. In this preliminary investigation, no attempt has been made to cover in great detail or accuracy any specific mission. At this state general parametric trends are of more importance than design numbers which must be related to specific configurations. However, the numbers presented may be useful for initial communications system evaluation for future missions.

6.2 REFINEMENTS IN THE THEORETICAL MODELS

The limitations of this study and those areas where improvement is possible are indicated in this section.

6.2.1 Trajectory

The calculation of the entry trajectories was based on the following assumptions:

- Exponential atmosphere
- Nonrotating planet and atmosphere
- Spherically symmetric planet
- Constant C_D and L/D for a given trajectory
- Non-ablating vehicle, i.e., constant mass and area.

These approximations are certainly reasonable for the present purpose. While refinements could be introduced in the last two items, they would result in a significantly more complicated computational procedure. This would not appear to be justified, particularly in view of the uncertainties in available estimates of the scale height to be used in the exponential atmosphere.

There is a great deal of uncertainty in the available data on the variation of pressure and temperature in the several planetary atmospheres. The electron density for a given case is dependent on the free stream pressure and temperature. Further, all the aerodynamic

parameters as well as the electron density will depend on how accurately the composition of the planetary atmosphere is known. These points are discussed in more detail in connection with the flow field model.

6.2.2 Flow Field

The flow field analysis is based on the following assumptions:

- The streamtube approximation assumes equilibrium chemistry, inviscid flow field, zero angle of attack, axisymmetric body, and pressure distribution on the nose of the body obtained from Newtonian theory.

- The boundary layer on the body is neglected since the region of interest is very close to the stagnation region. Only the inviscid shock layer outside the boundary layer is investigated, as noted below.

- An axisymmetric body shape is considered. The surface is noncatalytic and uncooled and the material is not ablating. Catalytic bodies will allow the atomic species to recombine at the surface and this may tend to decrease the electron density. A cooled body would also decrease the electron concentration; however, if the cooling is due to ablation, the ablating material may change the electron density significantly. For example, if the ablative material becomes ionized in the shock layer and produces its own electrons, there may be a net increase in electron density.

- The effect of turbulence has been ignored. Since the general subject of chemically reacting turbulent flow is not clearly understood, it is very difficult to predict the effect of this assumption on the plasma properties.

- The shock layer is assumed to be inviscid and the chemical state is in equilibrium. These assumptions are valid over a relatively limited Mach number-altitude range. The simplified order of magnitude analysis of Probstein[11] was used to estimate these regimes for entry into Venus and Mars, as shown in Figures 6-16 and 6-17.

- Thermal radiation from the hot gases in the shock layer has not been included in the parametric analysis. In general, the gas radiation will cause a decrease in the electron density due to the lower temperatures of the shock layer.

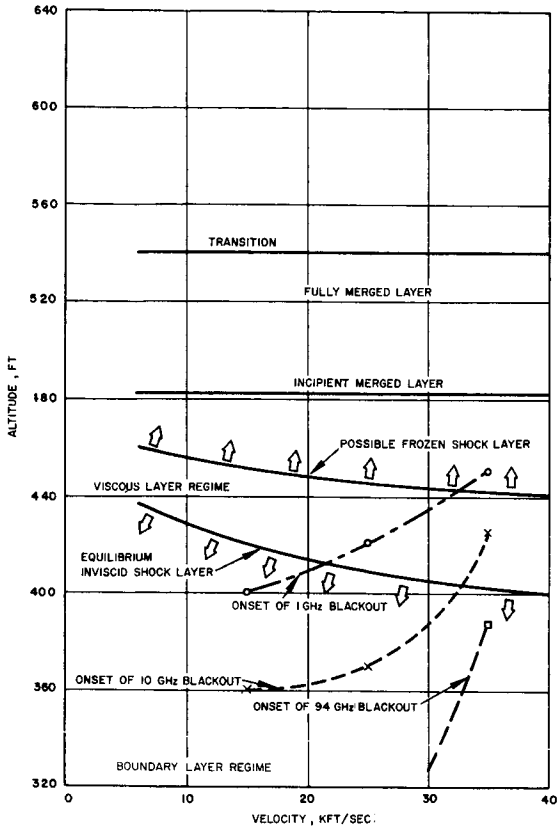


Figure 6-16. Stagnation region flow regimes for entry into a Venusian atmosphere by the hemispheric type vehicle. Blackout data is shown for the forward antenna.

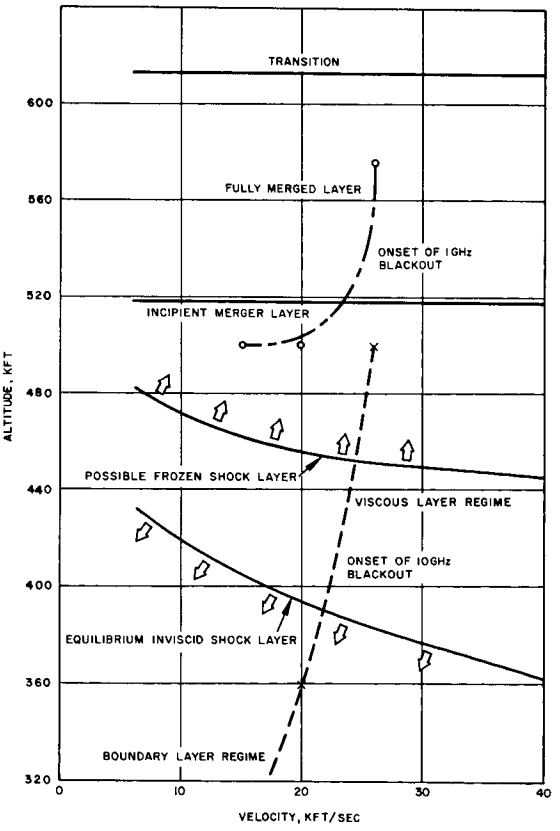


Figure 6-17. Stagnation region flow regimes for entry into a Martian atmosphere (5 mb surface pressure) by the Voyager type vehicle. Blackout data is shown for the forward antenna.

Figures 6-16 and 6-17 represent the flow regimes at the stagnation point of a highly cooled hemispherical body. However, the present configurations are neither completely hemispherical nor are they highly cooled. Also, the flow field analysis is not restricted to the stagnation point region. Nevertheless, the results will indicate the desired trend since the probes are at least hemispherically blunted and the flow field analysis is extended not too far from the stagnation region (up to 80^{o}). The lines separating the boundary layer flow regime will tend to shift a little for the uncooled case. The absence of cooling in the analysis provides a conservative estimate of the electron density, in any event, since it is quite likely that the actual probe will be cooled.

The assumption of equilibrium inviscid shock layer at the stagnation region will be valid in the lower altitude range (Boundary Layer Regime) as indicated by the arrows. At higher altitudes, in the Viscous Layer Regime, due to the increased rarefaction the shock layer can no longer be considered inviscid and the chemical state probably will be characterized by nonequilibrium effects. The present flow field analysis will not be applicable in this region as well as the other regions at higher altitudes. It is quite likely that the electron density in the Viscous Layer Regime will be greater than the equilibrium value due to the higher temperatures associated with the nonequilibrium effects in the shock layer. It is extremely difficult to estimate the extent of the increase in electron density without carrying out a nonequilibrium analysis. Such an analysis is beyond the scope of the present program.

With higher altitudes and more rarefaction the chemical state of the shock layer will be nearly frozen and the electron density will be fairly close to the freestream value. The Merged Layer Regimes and the Transition Regime are characterized by such frozen shock layers. Thus for a probe entering into a planetary atmosphere the blackout problem will probably first arise in the Viscous Layer Regime, and it may extend through a major part of the Boundary Layer Regime. The

altitude range in which such a problem arises, depends on the size of the probe, its velocity and the characteristics of the atmosphere through which the probe travels.

Figure 6-16 shows the flow regimes at the stagnation region of a 156 cm radius hemispherical body entering the atmosphere of Venus. The atmosphere is the mean model for Venus given by NASA TM X-53273, 1965. Figure 6-17 is drawn for a 200 cm radius body entering Mars (5 mb atmosphere given in Mariner IV data).

The trajectories are at constant velocity for the altitudes of interest and are in fact dependent only on the entry velocity. The altitude at which blackout begins in shown for the three frequencies in Figures 6-16 and 17. Note that the 100 GHz calculations are not affected by the gasdynamic considerations. The worse case is shown in Figure 6-17 for 1 Ghz. The significance of this effect is that the onset of blackout at 26,000 fps, for example, occurs in the vicinity of 440,000 feet altitude, rather than 575,000 feet, and the blackout time is reduced by five seconds. In general, therefore, the correct fluid dynamic model may cause the upper altitude limit to be lowered substantially, as expected, however, the total blackout time will not be significantly changed.

6.2.3 EM Wave Propagation

Signal loss calculations are based on the classical idealized model [12,13] for the propagation of plane waves through stratified media. These assumptions are marginal in several respects for the present applications. The wave may not be plane, either due to the type and location of the source relative to the plasma, or due to the curvature of the boundaries or transverse variations in the plasma properties. General solutions are not available to assess the adequacy of the model. A full understanding of the effect of the plasma on antenna performance, particularly in the near field, is not available. This includes antenna impedance mismatch and pattern distortion, leaky waves, and detailed near-field factors.

6.3 RECOMMENDATIONS FOR FUTURE ANALYSIS

As previously remarked, this study has purposely been made broad, at the expense of detailed design data. The purpose has been to establish the general effects of the parameters and to clarify the areas for future analyses and research. These recommendations are discussed below under the general categories of technology and systems studies. It should be remarked that any refined analysis of planetary entry communications blackout must include very specific design information. Particular characteristics of geometry, weight, and antenna location affect blackout significantly. The same is true for the entry or flight parameters so that detailed mission analyses must be made. It is virtually impossible to extrapolate overall theoretical or experimental results to account for small changes in conditions without careful consideration of the appropriate parameters involved.

6.3.1 Technology Studies

Our knowledge of the various atmospheric models is increasing steadily. Although it should not be necessary to modify the trajectory formulation used herein, this information is vital for the fluid dynamic considerations. The study of planetary atmospheres is outside the scope of this work so that the uncertainties involved must be treated in an appropriate parametric manner.

The flow field analysis must be made for specific geometry, as noted above. A continuum analysis is undoubtedly satisfactory although considerable work remains to be done in both the boundary layer and viscous layer regimes. (See Figures 6-16 and 6-17. Several of the topics which require considerable study are: flow fields about actual bodies including angle of attack; nonequilibrium, radiation, and ablation effects; and turbulence.

Interactions between the plasma and the communications system itself must be studied in greater detail. A major analytic problem remains before predictions of wave propagation through the plasma sheath can be considered reliable. A general description of the

non-planar wave interaction with non-stratified inhomogeneous plasma in the antenna's near field, a most formidible problem, is being worked on but as yet is not within the state of the art. Because of this, two important effects cannot be evaluated: the changes induced by the sheath both on the antenna's input impedence and on the radiation pattern. Depending on many factors these effects can have either little or a very great influence on the blackout time. The inclusion of the true non-plane wave and non-stratified plasma into the propagation calculation must therefore be made for many of the vehicular shapes and antenna locations expected. Meaningful laboratory and flight data would be invaluable.

6.3.2 Systems Study of Blackout

The above discussion has dealt with improving the analytic capability to predict blackout. Some mention should be made here of the possibility of alleviating it. Many schemes have been proposed which, in theory, will work but as yet none have been made practical for the blunted vehicle (the only type presently being considered for interplanetary missions). It appears unlikely that in the next decade there will be any breakthrough which will provide a 100 percent cure. There is hope, however, that by an optimum selection of all the parameters mentioned here, as well as others such as antenna polarization and angular orientation relative to the sheath, [12, 14] that the blackout can be minimized. The present study is a first attempt at looking at this overall system approach to the blackout problem.

6.4 REFERENCES

1. P.W. Huber, and T.E. Sims, The Entry-Communications Problem, Astronautics and Aeronautics, Vol. 2, No. 10, 1964, 30.

2. T.E. Sims, Communication Through Reentry Plasma, 1966 Winter Convention on Aerospace and Electronic System Conference Record, Vol. 1, IB-21.

3. R.R. Gold, Further Analysis of Propagation of Plasma Waves in a 'Spokewheel' Magnetic Field, Journal of Res. NBS Vol. 69D (Radio Science) No. 9, 1965, 1227.

4. J. Hoffman, Control of Plasma Collision Frequency for Alleviation of Signal Degradation, The Mitre Corporation, TM-04097, USAF No. ESD-TR-65-91, AD 467476, July 1965.

5. M.J. Kofoid, Permantly Magnetized Ferrite Antenna Windows For Improving Electromagnetic Wave Transmission Through a Plasma, Boeing Scientific Research Lab., DI-82-0442, AD618381, June 1965.

6. Astronautics and Aeronautics - 1985, Astronautics and Aeronautics, Nov. 1965.

7. Avco/Rad, Comparative Studies of Conceptual Design and Qualification Procedures For a Mars Probe/Lander, Part I Report, Vol. 3 and 4, RAD-TR-65-29, 8 October 1965.

8. I. Stambler, Mariner/Voyager: A New Fix on Mars, Space/Aeronautics, May 1966.

9. J.H. Allen and A.J. Eggers, A Study of the Motion and Aerodynamic Heating of Ballistic Missiles Entering the Earth's Atmosphere at High Supersonic Speeds, NACA Technical Report 1381, 1958.

10. P. Adams, High-Temperature Equilibrium Thermodynamic Properties of Chemically Reacting Mixture of N_2, O_2, CO_2, A, and Ne, Aerospace Technology Laboratory Research Report No. 2, Hughes Aircraft Co., November 1965.

11. R.F. Probstein, Shock Wave and Flow Field Development in Hypersonic Reentry, ARS Journal 31, 185-193, 1961.

12. K.G. Budden, Radio Waves in the Ionosphere, Cambridge University Press, 1961, Chapter 8.

13. R. R. Gold and R. Mason, Electromagnetic Wave Propagation Through Magnetoactive Plasmas, Report No. TDR-69(2119)TR-3, Aerospace Corp., Laboratory Div., February 1962.

14. R. R. Gold and J. F. Cashen, Communication Through Plasmas Bounding Slender Entry Vehicles, AIAA Paper No. 66-481, 4th Aerospace Sciences Meeting, Los Angeles, Calif., June 27-29, 1966.

7.0 RESEARCH SUBJECTS

There are a number of areas, both general and specific, where research and development effort will be required if deep space communications systems are to approach the full potential performance as delimited by fundamental considerations. For the most part the need for research in these areas has been mentioned or briefly discussed in pertinent sections of this report with reference to specific technological limitations. A comprehensive list of these research subjects is presented in summary form in Sections 7.1 through 7.5. Subjects of special relevance to the definition and development of a deep space communication system are indicated by an asterisk. These subjects are then regrouped in Section 7.6 into several programs related to the three candidate system configurations.

7.1 SYSTEM CONSIDERATIONS

*Evaluation of Candidate System Configurations (Ref. Section 3.4)

Realistic economic evaluations of the various candidate system configurations based on specific system designs should be made. These include studies of:

- Direct signal transmission from lander versus relay through orbiting bus[1]
- Direct optical signal return to multiple ground-based stations (for weather diversity) versus relay through Earth satellite station
- Direct microwave signal return to ground station versus preferred optical configuration

*A Computer Program for Evaluating R & D Expenditures (Ref. Section 7.6)

A computer program is needed to determine the optimum distribution among the various system parameters requiring advancement.

*Vehicle Control and Stability Requirements (Ref. Section 5.3)

A consideration need be made of conflicting requirements on vehicle control and stability among the various mission functions of and methods for their resolution. These include time-sharing of vehicle control and the use of an ancillary vehicle to meet the stringent requirements for stability of an optical communication system.

*Relation Between Transmitter Beamwidth and Point Accuracy (Ref. Section 5.3)

An analysis is required to determine the optimum relationship between beamwidth and pointing accuracy based on maintenance of a given average error rate and minimization of transmitted power as criteria.

Bit Error Rates in Photon Limited Detection (Ref. Appendix F)

An experimental verification of bit error rates should be made for various coding schemes for both coherent and noncoherent detection as the signal level approaches the photon limit. (This is discussed further in Appendix F).

Signal Correlation Techniques (Ref. Section 2.5)

An investigation of new signal processing techniques (such as spatial filtering or associative processing) should be made for the correlation of signals arriving with random phase due, for example, to atmospheric distortion effects.

*Information Correlation Techniques (Ref. Appendix E)

An investigation of signal processing techniques (including analog and digital time delays and associative processing) need be made for the correlation of wide-band signals collected over spatial dimensions large with respect to c/B, the propagation velocity over the bandwidth. (This is discussed further in Appendix E).

Comparison of Coherent versus Noncoherent Optical Systems (Ref. Section 5.4)

A detailed tradeoff study should be made between coherent and noncoherent optical communication systems (in optimum configurations) with reference to system complexity, reliability, and overall cost.

7.2 RADIO FREQUENCY TECHNIQUES AND COMPONENTS

Analysis of Noise Correlation in Subapertures (Ref. Section 4.5.5)

Analysis has been carried out under the assumption that the noise due to the sun in the phase-locked loops is statistically independent although in actual practice there will be some correlation from element to element. To assess the effect of the sun noise accurately, the cross-correlation of the noise from element to element should be taken into account.

*Optimum Size and Arrangement of Antenna Array Elements (Ref. Section 4.5.5)

Optimum design considerations should be studied with reference to minimum signal-to-noise ratio for lock-on, atmospheric and signal correlation lengths, scanning angle, and suppression of grating lobes.

*Phase Correlation and Multiaperture Arrays (Ref. Section 4.5.5)

Development of techniques for correlation of the phase among the various elements of a multiaperture array is required. These include self-phasing techniques to provide self-steering and compensation for atmospheric distortions. (See also Section 7.1, Signal Correlation Techniques.)

*Sub Aperture Sidelobe Control (Ref. Section 4.5.5)

Two alternatives should be investigated for subaperture sidelobe control: the first alternative is to provide a suitable illumination taper across the subaperture, sufficient to give the desired sidelobe characteristics; the second method is to employ a combination of subaperture illumination taper and illumination taper across the composite aperture.

*Time-Delay Correlation over Large Apertures (Ref. Appendix E)

(See Section 7.1, Information Correlation Techniques.)

Subaperture Time Delay Compensation (Ref. Appendix E)

Detailed analysis will be required to determine the extent of garble and modulation distortion introduced as a function of array element dimension due to the distance travelled by an electromagnetic wave during the bit duration.

Initial Phasing of Subapertures Using Statistical Techniques (Ref. Appendix E)

Statistical techniques of initial phase correlation such as introducing random phases into the signal line from each subaperture, require additional study to determine the optimum method in terms of cost, speed and effectiveness.

*The Use of a Computer for Control of a Large Ground Array (Ref. Appendix E)

The design of a special purpose computer or, possibly, use of the capabilities inherent in a large general purpose computer for control of the entire multiaperture receiving system and for signal processing warrants considerable additional study.

Precision Angle Tracking with a Multi-Element Array (Ref. Appendix E)

A detailed study should be made of precision angle tracking by utilizing the various subapertures of a large array in either a phase monopulse or in an interferometer configuration, taking into account the specific characteristics and detailed design of the electronically-scanned subapertures.

*Evaluation of Discrete Array Elements versus Subarrays (Ref. Section 4.5.5)

A tradeoff study should be made with reference to performance and cost of discrete array elements versus subarrays comprising iterative subelements.

*Antenna Feed and Scanning Techniques (Ref. Section 4. 5. 5)

An investigation and comparison of mechanical, electrical, and hybrid methods is required for scanning of large ground arrays over specified angular coverage.

*Electronic Phase Shifters (Ref. Section 4. 5. 5)

Consideration should be given to an optimization of electronic phase shifter characteristics by improvement of parameters such as RF loss and switching power at a sacrifice in switching time. In the cases of diode, ferrite, and plasma shifters, further investigation should be carried out in anticipation of future lower costs and improvements in diodes and ferrite materials.

*Subarray Design (Ref. Section 4. 5. 5)

Development of electrical design and fabrication techniques for the production of low-cost subarrays is needed.

*Reflectarray Antennas (Ref. Section 4. 5. 5)

Due to their inherent simplicity, especially for large ground arrays, intensive design investigations of the application of the reflectarray technique should be initiated.

*Evaluation of Spacecraft Antenna Configurations (Ref. Section 4. 5. 4)

A comparison is needed with reference to performance and burden of the various approaches to spacecraft antenna design, including simple antennas, discrete inflatable or erectable antennas, and multielement phased arrays.

*Self-Steering Array Techniques for Spacecraft Antennas (Ref. Section 4. 5. 4)

Development of self-steering array techniques applicable to lightweight multiaperture, distributed drive, spaceborne antennas is needed.

*Phased-array Components for Spacecraft Antennas (Ref. Section 4.4)

Development of efficient, lightweight RF amplifiers or their equivalent such as phase-locked solid-state oscillators is a critical need.

Millimeter and Submillimeter Power Sources (Ref. Section 4.1)

Research and development of millimeter- and submillimeter-wave power sources is needed. These include solid-state transit-time devices and harmonic generation techniques.

Millimeter and Submillimeter Detectors (Ref. Section 4.2)

Research and development of millimeter- and submillimeter-wave detectors including balanced mixers is needed for frequencies above 70 Ghz and for hot-carrier semi-conductor detectors[2].

Low-Noise Preamplifiers (Ref. Section 4.3)

Research should be directed toward cost reduction and mass production techniques on cooled parametric amplifiers, and a search made for a tunnel diode amplifier operating at cryogenic temperatures with noise performance approaching that of the cooled paramp.

7.3 OPTICAL TECHNIQUES AND COMPONENTS

*Basic Laser Mechanisms (Ref. Section 5.1)

Research into the detail mechanisms of existing lasers with particular attention to the CO_2 and other vibrational-level gas lasers is needed. This is to be aimed at the discovery of improved energy conversion mechanisms or of new laser materials.

*New Laser Configurations (Ref. Section 5.1)

Research into new laser configurations (e.g., ring discharge, folded lasers) and excitation mechanisms for potential reduction in size and weight and increase in efficiency and reliability of lasers is needed.

Laser Thermal Control (Ref. Section 5.1)

An investigation of methods for the exclusion and removal of waste energy from the laser cavity in order to reduce undesirable thermal effects and increase reliability is needed.

Laser Pumping (Ref. Section 5.1)

Development of techniques to restrict the generation of pump energy to the effective laser pump bands and thus improve efficiency, minimize thermal effects, and extend lifetime is needed.

Sun-Pumping of Laser[3,4] (Ref. Section 5.1)

Development of techniques for the collection, and concentration in the laser cavity, of solar radiation within the laser pump bands of interest, and of methods for better matching of the laser pump bands to the solar spectrum could be investigated.

Laser Frequency Doubling (Ref. Section 5.1)

An investigation of frequency doubling techniques could be made with particular reference to doubling of the neodymium laser line from 1.06 to 0.53 microns in order to provide a near optimum spectral match to the S-20 photosurface.

*Laser Frequency Control (Ref. Section 5.1)

An investigation and development of methods for stabilization and tuning of laser frequencies including the generation of harmonic and non-harmonic frequencies (e.g., utilizing parametric techniques[5]) is needed.

*Optical Modulators (Ref. Section 5.2)

Research on non-linear optical materials (particularly for 10.6 microns) and on slow-wave structures suitable for the interaction of the optical and modulating fields is needed.

*Intracavity Modulation Techniques (Ref. Section 5.2)

Research and development of techniques for optical modulation within the laser cavity to obtain higher modulation efficiencies associated with multiple passes of the beam are needed.

*Solid-State Photodetectors (Ref. Section 5.5)

A comprehensive experimental and theoretical program to determine and improve the characteristics (quantum efficiency, internal gain, impedance level, response time, operating conditions) of solid-state detectors, with particular reference to the far infrared region of the optical spectrum is needed.

*Coherent Optical Receiver Techniques (Ref. Section 5.4.1)

Development in general, of coherent receiver techniques with emphasis on frequency and phase control of local oscillator and on methods for increasing the effective field of view (e.g., detector arrays and electronic tracking) in order to alleviate aperture restrictions imposed by pointing errors is needed. (See also below.)

*Electro-Optical Stabilization, Pointing, and Tracking (Ref. Section 5.3)

An investigation of electro-optical techniques for fine pointing of an optical transmitter in reference to a received (beacon) signal with emphasis on the problems of alignment, lead angle, and self-calibration is needed.

Electronically Steerable Optical Antenna (Ref. Section 5.3)

An investigation of methods for inertialess steering of an optical transmitter or receiver beam over angular dimensions of the order of a degree or more is needed.

*Optical Phased Array Techniques (Ref. Section 5.4.1)

A consideration of methods and techniques for phase correlation of elements in an optical antenna array including self-phasing techniques to counteract atmospheric distortion effects is needed. (See also Signal Correlation Techniques under Section 7.1.)

*Large Coherent-Light Collectors (Ref. Section 5.4.1)

A study is needed of means, such as Fresnel-zone plates, for exploiting the monochromaticity of the laser signal to make possible large, light-weight apertures having loose dimensional tolerance requirements.

Large Noncoherent-Light Collectors (Ref. Section 5.4.2)

A study of the practical limitations imposed on the fabrication and use of large light-collecting apertures ("photon buckets") for noncoherent receivers is needed.

7.4 ATMOSPHERIC STUDIES

Atmospheric Transmission (Ref. Section 2.4)

Measurements of atmospheric transmission as a function of zenith angle are needed to determine more quantitatively:

- Attenuation for a variety of weather conditions in the millimeter-wave region of the spectrum,
- Attenuation in the submillimeter-wave region and the extreme infrared,
- Attenuation for a variety of weather conditions in the infrared and visible regions,
- Absorption fine structure in the infrared and visible particularly in the vicinity of laser wavelengths of interest (e.g., 0.50, 0.53, 0.633, 0.694[6], 0.84, 1.06, 2.36, 3.5, 9.0, 10.6 microns).

Atmospheric Scattering (Ref. Section 2.4)

Measurement of the angular distribution of scattered radiation is needed as a function of meteorological conditions, altitude, and frequency in:

- The micro- and millimeter-wave regions of the spectrum,
- The infrared and visible regions.

Analysis of multiple scattering effects at optical wavelengths[7] is also of interest.

*Atmospheric Distortion (Ref. Section 2.5)

Measurements of wavefront deviations and quantitative determination of atmospheric correlation lengths as a function of weather conditions, zenith angle, and frequency are needed:

- In the micro- and millimeter-wave portions of the spectrum[8],

- In the infrared (especially at 10.6 microns) and visible regions[9, 10] (e.g., a multiaperture coherent receiver providing real-time data on variations in phase and amplitude over the wavefront from a monochromatic source).

7.5 PLASMA PROPAGATION

Methods for Alleviation of Communication Blackout (Ref. Section 6.3)

A comparative study of the several existing proposals for alleviating communication blackout is needed. These include:

- Optimum design considerations for antenna location, orientation, and polarization; signal power and frequency; and aerodynamic shaping of vehicle;
- Material additives to reduce electron concentration;
- "Window" effect derived from application of a magnetic field;
- The use of optical frequencies.

Magnetoactive Effect (Ref. Section 6.3)

A detailed analysis of the use of magnetic fields for the elimination of blackout, including general systems and engineering considerations is needed.

Communication Blackout during Blast-Off (Ref. Section 6.3)

An examination of the blackout problem during blastoff as it differs from the problem during atmospheric entry is needed.

7.6 RECOMMENDED RESEARCH PROGRAMS

The primary objective of the Advanced Deep Space Communication Systems Study has been to define those areas of technology where research effort can best be expended in order to meet deep space communication needs in the period 1970 to 1980. As a result of this study three primary system configurations have been selected as most promising. These systems are:

- a direct microwave spacecraft-to-earth communication link in the region of 1 to 5 GHz;
- a satellite relay configuration (one or at most two required) utilizing a 10.6 micron optical link from spacecraft to satellite with a noncritical microwave link for the short range from satellite to earth;
- a direct optical link at 10.6 microns employing additional ground stations to assure the necessary weather diversity.

It is the purpose of this section to re-state, in an ordered format, those research projects which are necessary for further system definition and for a realistic choice among the three candidates.

7.6.1 Criteria for System Selection

While the present study has used data quantity and quality as criteria for selection of the candidate systems, taking only an implicit account of costs, the final choice must included explicit cost considerations. Further analysis is required to establish costs on a quantitative level. This will require more-detailed system designs including cost considerations of the various components which the three candidate systems comprise. Pertinent individual study topics (taken from section 7.1 through 7.4) have been organized below according to the three configurations.

As a guide to the establishment of design goals for the individual system components a computer program, which is a modification of an existing program, is suggested. The existing program was developed under NASA contract NAS 5-9637 between the Hughes Aircraft Company and NASA-Goddard Space Flight Center. The computer program

determines the optimum split of burden (cost or weight) in order to achieve maximum deep space communication performance for a given burden. Parameter values are calculated for the transmission equation at a given range as a function of bit rate. As the bit rate increases the parameter values increase until a preset parameter "stop" is encountered, e.g. the maximum diameter allowed for a transmitting antenna. The program calculates the means of achieving higher bit rates by increasing the remaining parameters at a rate faster than that prior to the parameter value "stop". This causes the total burden to increase at a rate faster than before the stop. A "stop" represents a technological limit, a limit which could be exceeded if research and development funds were to be expended.

The basic programming of the computer has been completed. The remaining effort would be a modification to provide quantitative guidance for research fund expenditures. The key comparison to be made is that of comparing the total burden of a given system with stops, to the total burden (including R & D funds) of a system in which R & D funds have improved the state of the art and which thus has no stops.

The results of this program can be used to provide a quantitative basis for apportionment of the increased system performance required among the various system parameters and to weigh the effort to be expended on the study topics as listed below in outline form.

7.6.2 RF System, (1-5 GHz)

There are two major ways of implementing the high data rate RF System (10^6 bits per second). One would utilize a high gain spacecraft antenna (45 to 50 db) with a single element receiving antenna having a gain of 60-65db. The other antenna configuration would use a 35 to 40 db antenna in the spacecraft with a 70 to 75 db antenna on the ground. Therefore additional study is required for both high gain spacecraft antennas and high gain ground arrays. The study topics of particular concern are summarized below.

- Comparison of various high-gain extensible spacecraft antennas with lower-gain self-steering and phased-array antennas.

- Study of ground antenna arrays including
  - Optimum size and arrangement of antenna array elements
  - Sub aperture sidelobe control
  - Time delay correlation over large apertures
  - The use of a computer for control of a large gound array
  - Antenna feed and scan techniques
  - Evaluation of descrete array elements versus subarrays
  - Subarray design.

7.6.3 Optical Satellite-Relay System (10.6 microns)

Optical communications techniques need further study in order to develop reliable components and to establish realistic cost and weight burdens.

For the Optical Satellite-Relay System the following topics need consideration.

- Coherent optical receiver design
- Electro-optical stabilization, pointing and tracking
- Solid state photo-detectors
- Laser techniques
  - Basic mechanisms
  - New configurations
  - Frequency control
- Optical modulators and intracavity modulation techniques
- Acquisition and tracking
  - Vehicle control and stability requirements
  - Relation between transmitter beamwidth and pointing accuracy

7.6.4 Direct Optical System (10.6 microns)

For the direct optical system the study topics listed in 7.6.3 apply with the following additions:

- Correlation of signals from multiple optical apertures
- Atmospheric distortion

7.7 REFERENCES

1. T.A. Barber, J.M. Billy, and R.D. Bourke, Systems Comparison of Direct and Relay Link Data Return Modes for an Advanced Planetary Mission, JPL Tech. Memo. 33-228, Jan. 15, 1966.

2. L.S. Stokes and K.L. Brinkman, Reference Data for Advanced Space Communication and Tracking Systems, Report P66-16, Contract No. NAS 5-9637, Hughes Aircraft Company ASG, Culver City, California, 1966.

3. G.R. Simpson, Continuous Sun-Pumped Room Temperature Glass Laser Operation, Applied Optics, 3, 6, pp. 783-784 (1964).

4. P.H. Keck, J.J. Redmann, C.E. White, and R.E. DeKinder, Jr., A New Condenser for a Sun-Powered Continuous Laser, Applied Optics, 2, 8, pp. 827 to 831 (1963).

5. J.A. Giodmaine and R.C. Miller, Tunable Coherent Parametric Oscillation in $LiNbO_3$ at Optical Frequencies, Phys. Rev. Letters 14, 24, 973 (1965).

6. R. K. Long and C.H. Boehnker, Measured Atmospheric Absorption at Ruby Optical Maser Wavelengths, Report 1641-10, Ohio State University, Research Foundation, Contract No. AF 33(657)-11195.

7. Discussion between Dr. M.R. Nagel (ERC) and Dr. S.J. Klapman (HAC) at NASA ERC, 1 March 1966.

8. J.E. Adler, et al., The Design of an Experiment to Determine the Limitations Imposed on a Multiple Aperture Antenna System by Propagation Phenomena, 3rd Quarterly Report, Contract No. NAS 5-3974, SRI Project 5067 Stanford Research Institute, Menlo Park, California, March 1965.

9. Laser Space Communications Systems (LACE) Study, North American Aviation, Inc., SID, NASA Contract NASw-977.

10. Proposal to NASA ERC, No. 66M-7102/B0581, Hughes Research Laboratories, 1966.

APPENDIX A
SUN IN THE SIDELOBES

The contribution to the antenna temperature from noise inputs through the pattern due to the temperature of the sun is given by

$$T_a = \frac{1}{4\pi} \int_{\Omega s} g T_s \, d\Omega_s \qquad \text{(A-1)}$$

where

T_a = antenna temperature

g = antenna gain in that portion of its pattern occupied by the sun

T_s = temperature of sun

Ω_s = solid angle of sun

This expression is completely general and gives the antenna temperature regardless of what portion of the pattern is occupied by the sun. It can be simplified for the purpose of studying effects of noise through the sidelobes as follows: The assumption is made that the gain of the pattern over the whole sun is equal to the gain of the envelope of the sidelobes at the edge of the sun nearest the main beam and this gain is designated g_o'. (See Figure 4-10.) This brings the variable out from under the integral sign and the integration becomes simply the solid angle subtended by the sun, which is given by

$$\Omega_s = \frac{\pi}{4}\left(\theta_s\right)^2$$

where

θ_s = angle subtended by the sun in radians

Then Equation (A-1) becomes

$$T' = \frac{1}{4\pi} g_o' T_s \left(\frac{\pi}{4}\theta_s^{\,2}\right) = \frac{\theta_s^{\,2}}{16} T_s g_o' \qquad \text{(A-2)}$$

where T' is always greater than T_a. To obtain an expression for T' in terms customarily used to describe antenna patterns, let

$$g_o = \frac{g_o'}{G_o} \qquad \text{(A-3)}$$

where

G_o = peak gain of antenna

g_o = ratio of gain in the sidelobe region compared to the peak gain

In this case g_o' is restricted to the envelope of the sidelobes and then g_o is the level of the envelope of the sidelobes normalized to a peak gain of unity. The sidelobe level is ordinarily designated as being so many db below the peak of the main beam and would be expressed in decibels as

$$g_{o_{db}} = g_o'_{db} - G_{o_{db}}$$

Since g_o' is less than G_o (and ordinarily much smaller), g_o will take on values from perhaps -20 db to a very large negative value, depending on the size of the antenna, the distribution across the aperture, and the magnitude of the angle between the main beam and the sun.
Rewriting Equation (A-3) as

$$g_o' = g_o G_o$$

and substituting into Equation (A-2) gives

$$T' = \frac{\theta_s^2}{16} T_s g_o G_o \qquad \text{(A-4)}$$

For large antennas (diameters of 1000 to 2000 feet), the 3-db width of the mainbeam is about 1 or 2 minutes of arc and, in the direction away from the main beam, the peaks of the far out sidelobes ideally diminish monotonically over the sun's diameter of 30 minutes of arc.

Thus, the sidelobe level integrated or averaged over the sun's disc would be smaller than g_o and consequently as indicated above $T_a < T'$. (See Figure 4-10.)

Since T' was set at 25°K, Equation (A-4) can be used to evaluate g_o. This has been done for several different values of G_o and the results are tabulated below. A gain, G_o, of 81.2 db at 2.3 Ghz corresponds to a large circular aperture, 2100 feet in diameter, with an efficiency, η, of 55 percent.

G_o	g_o
81.2 db	-64.2 db
80	-63
70	-53
60	-43
50	-33

Table A-1. Evaluation of g_o for Different Values of G_o

As an illustration of a technique that can be used to determine the angular distance of closest approach that the sun may take to the main beam without exceeding an antenna noise temperature of 25°K, large circular apertures with illumination functions of the form $(1-r^2)^p$ may be considered. The characteristics of this family of distributions are reproduced in Table A-2 from Silver* for convenience.

*See reference 24, Section 4.6.

p	Efficiency, η	Half-power Beamwidth, θ_a	Position of First Zero	First Sidelobe (db below peak)
0	1.0	1.02 λ/D	1.22 λ/D	17.6 db
1	0.75	1.27	1.63	24.6
2	0.56	1.47	2.03	30.6
3	0.44	1.65	2.42	36.0
4	0.36	1.81	2.79	40.9

Table A-2. Characteristics of Circular Apertures

The efficiencies of these distributions are related to the parameter p as follows:

$$\eta = \frac{1 + 2p}{(1 + p)^2} \tag{A-5}$$

The gain, G_o, is then calculated from the conventional formula

$$G_o = \frac{4\pi A \eta}{\lambda^2} = \left(\frac{\pi D}{\lambda}\right)^2 \eta \tag{A-6}$$

where

D = diameter of the aperture

λ = wavelength in the same units as D

The normalized patterns of these circular apertures are given by

$$g(u) = \left\{\frac{(p+1)!\ J_{p+1}(u)}{(\frac{u}{2})^{p+1}}\right\}^2 \tag{A-7}$$

where

$$u = \frac{\pi D}{\lambda} \sin\theta \tag{A-8}$$

and θ = angle off the mainbeam in radians

By using the asymptotic representation for the Bessel Function for large u and setting the cosine term equal to unity, an approximate expression for the envelope of the far out sidelobes may be developed:

$$J_{p+1}(u) < \frac{2}{\pi u}, \quad u > p + 3$$

and

$$g(u) < \left\{ \frac{(p+1)!}{\sqrt{\pi}} \left(\frac{2}{u}\right)^{p+\frac{3}{2}} \right\}^2 \tag{A-9}$$

Since the right-hand term is always greater than the peaks of the sidelobes in the normalized pattern, g(u), it is reasonable to equate the right side with g_o, the term previously defined as the ratio of the envelope of the sidelobe gain to peak gain. Then

$$g_o = \left\{ \frac{(p+1)!}{\sqrt{\pi}} \left(\frac{2}{u}\right)^{p+\frac{3}{2}} \right\}^2 \tag{A-10}$$

Solving Equation (A-10) for u gives

$$u = 2 \left[\frac{(p+1)!}{\sqrt{\pi\, g_o}} \right]^{\frac{2}{2p+3}} \tag{A-11}$$

Substituting for u from Equation (A-8) yields

$$\sin \theta_o = \left[\frac{(p+1)!}{\sqrt{\pi\, g_o}} \right]^{\frac{2}{2p+3}} \left(\frac{2\lambda}{\pi D} \right) \tag{A-12}$$

where θ_o is the particular angle off the mainbeam specified by the parameters of Equation (A-12).

Since the antennas under discussion are of very high gain, the sidelobe envelope will drop off to the necessary low level within a small angle. Therefore the small angle approximation for sin θ may be used (sin $\theta \simeq \theta$). Also it will be convenient to use G_o in Equation (A-12) instead of the $\frac{\lambda}{\pi D}$ term; hence, the gain formula, Equation (A-6) is rewritten as

$$\frac{\sqrt{\eta}}{\sqrt{G_o}} = \frac{\lambda}{\pi D} \tag{A-13}$$

Equation (A-12) then becomes

$$\theta_o = \left[\frac{(p+1)!}{\sqrt{\pi\, g_o}}\right]^{\frac{2}{2p+3}} \frac{2\sqrt{\eta}}{\sqrt{G_o}} \tag{A-14}$$

Substituting from Equation (A-5) for η gives

$$\theta_o = \left[\frac{(p+1)!}{\sqrt{\pi\, g_o}}\right]^{\frac{2}{2p+3}} \frac{2}{\sqrt{G_o}} \sqrt{\frac{1+2p}{1+p}} \tag{A-15}$$

With this expression it is possible to hold G_o constant (and hence its corresponding g_o) and study the behavior of θ_o as the distribution parameter, p, is varied. Increasing p results in lower efficiency, thereby requiring a larger aperture to maintain the fixed G_o. However, increasing p reduces the sidelobe level and, hence, the angle θ_o, which means that the main beam can look closer to the sun without exceeding a noise temperature of 25°K. This reduction in θ_o can further be interpreted as a reduction in the number of days that communications with the deep space probe are degraded due to excess noise from the sun entering through the near-in sidelobes or the main beam. A trade-off is indicated of increased cost for the antenna versus a reduction in "degraded time" for spacecraft to earth communications.

To facilitate this trade-off study, a normalized effective diameter of the sun is defined:

$$\rho = \frac{\text{Effective diameter of sun}}{\text{Optical diameter of sun}} \qquad \text{(A-16)}$$

Since the peak of the mainbeam of the antenna can come no closer to the edge of the sun than θ_o without have T' exceed 25°K, the effective radius of the sun can be said to be θ_o plus the optical radius of the sun. The effective diameter then becomes $2\,\theta_o$ plus the optical diameter. If the optical diameter of the sun is taken $1/2^o$ and θ_o is expressed in degrees, Equation (A-16) becomes

$$\rho = \frac{2\,\theta_o{}^o + \frac{1^o}{2}}{\frac{1^o}{2}} = 4\,\theta_o + 1 \qquad \text{(A-17)}$$

where ρ is a dimensionless quantity. For example, a value of 3 for ρ simply means that the peak of the main beam must avoid a circular area around the sun with a diameter three times the optical diameter of the sun to assure $T' \leq 25^oK$.

Trying to estimate the number of days that communications will be degraded with a typical deep space probe is a difficult task to accomplish with a high degree of accuracy. However, some information is known about trajectories that JPL is contemplating for early fly-by missions to Jupiter, and these will be used as a guide. These trajectories are all long in terms of time and range from approximately 1-1/2 to 2-1/2 years in duration. In each case, the sun occults the probe optically (or comes very near to occulting it) at least once, and in some cases as many as three times during a single mission. Twice will be assumed to be typical. Since the sun will have an effective diameter ρ times its optical diameter, the number of times that the probe is effectively occulted (and the duration of each such occultation) will be a function of ρ and the plane of the trajectory in reference to the ecliptic.

The duration of an optical occultation in which the center of the sun passes directly between the earth and probe can be estimated by assuming that the probe is far out beyond the earth's orbit and moving at a fairly slow rate in angular velocity. The velocity of the earth in its orbit is then the prime determining factor, and occultation will occur over an arc of the earth's orbit approximately equal to the arc of the sun's diameter as seen from earth. (See Figure A-1.) Similarly, the arc of the earth's orbit over which performance will be degraded will be approximately equal to the angle subtended by the effective diameter of the sun at the earth. Since the earth takes 365 days to traverse 360^{o} of arc it can, to a very close approximation, be said that an effective diameter of the sun of 1^{o} will cause an effective occultation time of 1 day. A sample estimate of the total degradation time for a typical mission as a function of ρ can be obtained by drawing two arbitrary paths across a diagram of the sun and appropriate circles representing successively larger normalized effective diameters and then graphically determining effective occultation time, N, for each value of ρ. This procedure was followed in Figure A-2.

The results of the integration are presented in Table A-3 below. This set of data is empirically approximated by Equation (A-18) which gives an indication of a more general form of the relation.

$$N = 0.187\,\rho^{2} - 0.15\rho + 0.213 \qquad \text{(A-18)}$$

ρ	N
1	0.00
2	0.75
3	1.37
4	2.62
5	4.13

Table A-3. Duration of Effective Occultation.

This equation was used to transform values of ρ calculated from Equation (A-17) into values of N. A more general relationship could be obtained by assuming some statistical distribution for the trajectory traces across Figure A-2 and integrating to determine N as a function of ρ. Time precluded an analysis of this type for this report. N is plotted in Figure 4-12 versus the first sidelobe level for antennas which have constant peak gains of 70 db and 80 db. It is apparent that a small amount of tapering produces a rapid initial decrease in N, but that the curve rapidly flattens out and successive amounts of tapering are progressively less profitable.

It is of some interest to extend the analysis to determine the level of the first sidelobe required to obtain an effective occultation time of one day as a function of peak antenna gain. Setting N = 1 in equation A-18 above, the value $\rho = 2.5$ is obtained. Referring to Figure 4-11, it can be seen that antennas with peak gains of 80 db, 70 db, and 60 db require first side levels of approximately 21 db, 25 db, and 41 db below the peak, respectively. A smooth curve drawn through these three points so determined is given in Figure A-3. The Figure serves to illustrate the approximate nature of the relationship between peak antenna gain and the first sidelobe level in db below the main beam for the effective occultation time chosen. It should be observed again that the curve flattens rapidly, and extremely low sidelobe levels are required as the antenna gain decreases.

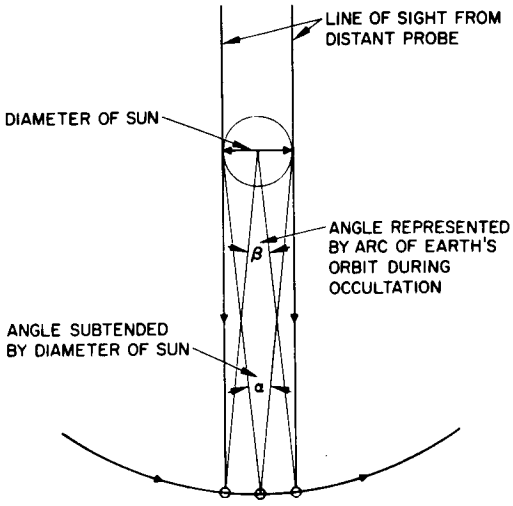


Figure A-1. Geometry relating diameter of sun to arc of earth's orbit during occultation.

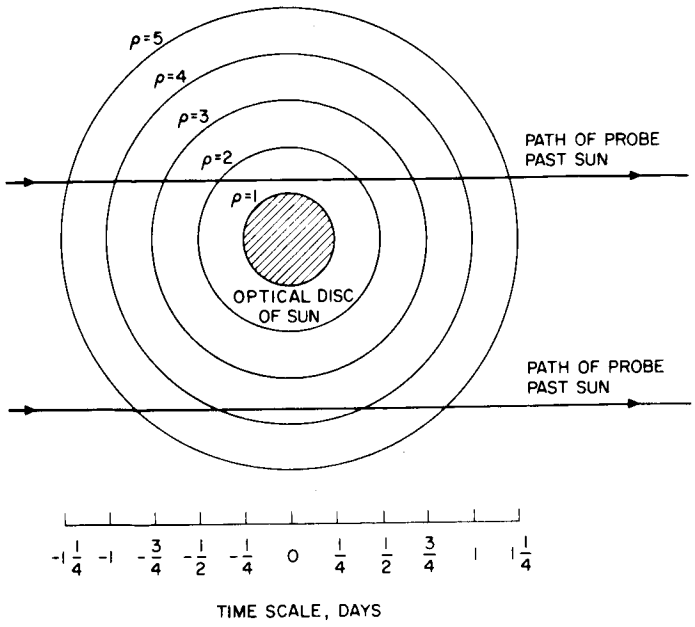


Figure A-2. Graphical determination of total effective occultation time for a typical deep space probe.

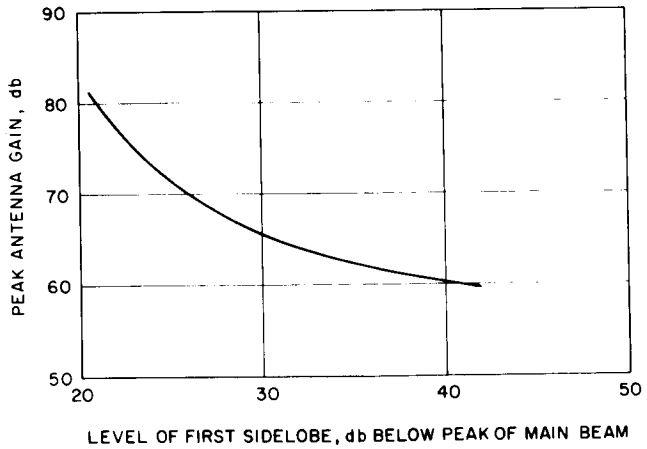


Figure A-3. Peak antenna gain versus first sidelobe level to achieve a one-day effective occultation time on a typical Jupiter trip.

APPENDIX B

COST ESTIMATE OF ANTENNA ARRAYS

The cost of a 70 and an 80-db gain earth station which consists of an array of large paraboloids has been estimated. The estimation is based on the experience that JPL has acquired in the operation of the NASA/JPL deep space network. Several assumptions were made to make this estimate. First, the gain of the array was assumed to be proportional to the total area of the paraboloids times an efficiency factor that would result from using a tapered illumination across the array. This condition should hold as long as the individual paraboloids are spaced far enough apart so that their mutual coupling is negligible. The second assumption made was that the paraboloids would be sufficiently randomly located in the aperture so that large grating lobes will not exist. It is expected that the sidelobe level at some angles will rise on a statistical basis because of this random location of elements but time has again precluded a detailed investigation of this effect. A third assumption made was that the array would not have to scan to large angles; hence four or more stations may be needed to cover the necessary solid angle.

The number of paraboloids required for the array is determined as follows. The area required for a specified gain is given by the gain formula rewritten as

$$A = \frac{G_o \lambda^2}{4 \pi \eta \eta_e} \qquad \text{(B-1)}$$

Where η_e, the element efficiency, must now be included.

The area of an individual array element is similarly related to its gain and efficiency by the expression:

$$A_e = \frac{G_e \lambda^2}{4 \pi \eta_e} \qquad \text{(B-2)}$$

The number of elements required, n, is obtained by dividing the area of the total aperture by the area of the element:

$$n = \frac{A}{A_e} \qquad \text{(B-3)}$$

which becomes upon substitution from Equations (B-1) and (B-2)

$$n = \frac{\frac{G_o \lambda^2}{4\pi\eta\eta_e}}{\frac{G_e \lambda^2}{4\pi \eta_e}} = \frac{G_o}{\eta G_e} \qquad \text{(B-4)}$$

The number of elements required is thus related to the peak gain of the array, G_o; the efficiency of the aperture distribution across the array, η; and the element gain, G_e. The numbers of elements required for a number of different situations have been determined. The results are tabulated in Table B-1.

The expression developed by Potter et al. * of JPL for station cost was used to estimate the cost of the ground antenna for the various parameters outlined in the table below. The expression was modified by the elimination of all operating costs, and the comparison is based on fabrication and installation costs only. (As with the JPL estimate – the cost of the land required is not considered.) Tables B-2 and B-3 are taken from the JPL report. *

p	First Sidelobe Level	η	G_o = 70 db		G_o = 80 db	
			Number of 85-foot Paraboloids Required	Number of 210-foot Paraboloids Required	Number of 85-foot Paraboloids Required	Number of 210-foot Paraboloids Required
0	17.6 db	1.0	20.8	7.9	208	79
1	24.6	0.75	27.8	10.6	278	106
2	30.6	0.56	37.2	14.2	372	142
3	36.0	0.44	47.4	18.0	474	180
4	40.9	0.36	57.9	22.0	579	221

Note: Fractional antennas are used for expediency in calculating costs.

Table B-1. Number of Elemental Paraboloids Required for 70 and 80 db Gain

* See Reference 10, Section 4.6.

Item	Master Cost		Slave Cost*	
	Nominal	Minimal	Nominal	Minimal
Facilities	$\$2.5 \times 10^6$	$\$2.5 \times 10^6$	$\$0.24 \times 10^6$	$\$0.12 \times 10^6$
Electronics	$\$3.1 \times 10^6$	$\$2.5 \times 10^6$	$\$1.80 \times 10^6$	$\$0.51 \times 10^6$
Operations	$\$2.6 \times 10^6$/yr	$\$2.6 \times 10^6$/yr	$\$0.61 \times 10^6$/yr	$\$0.22 \times 10^6$/yr

* Add a fixed array-controller cost of $\$0.5 \times 10^6$ for $n > 2$. A learning curve of 0.95 is applicable to the first slave and every doubling of the total number of antennas thereafter. (Applies to facilities and electronics.)

Table B-2. Selected costs

Resultant	Computation
STATION COST C =	(Cost of n antennas of Diam D $= n \times 0.95^{\log_2 n} \times 4.37D^{2.78}$) + (Cost of master electronics and facilities) + $(n - 1)\ 0.95^{\log_2 n}$ x (Cost of slave electronics and facilities)

Table B-3. Station costs

The cost of a single station as determined from the modified formula of Table B-3 and Table B-2 is plotted in Figure 4-13; two curves result as a function of first sidelobe level for each value of peak gain, one for an array of 85-foot paraboloids and one for an array of 210-foot paraboloids. It is apparent from this graph that cost for the 210-foot elements starts higher and goes up faster as the first sidelobe is lowered than it does for the 85-foot elements. It is also apparent that N, the number of days of lost communication time, can be reduced from 4 days to 1 day for an increase in cost of about 10 percent. However, beyond that point it rapidly becomes more and more expensive to reduce N, with the resulting conclusion that a taper across the array aperture that produces first sidelobe levels between -20 and -25 db is probably near optimum.

APPENDIX C

SCANNING DEGRADATION EFFECTS IN PLANAR ARRAYS

A large multi-wavelength planar array of identical elements located at points on a rectangular lattice structure lying in the x-y plane can be considered. The condition that the array be large implies that edge effects can be ignored. The interelement spacing in the x direction can be denoted by d_x, and that in the y direction d_y, and the array can be assumed to radiate into the half-space $z > 0$. All elements are independently and identically fed by circuits which consist of a transmission line of characteristic impedance, Z_o, a generator with internal impedance, Z_g, and a phase shifter that is matched to the generator impedance.

When the entire array is excited with a uniform or nearly uniform amplitude and a progressive phase shift, a net reflected wave will, in general, exist in each transmission line. This net reflected wave is generated from the mismatch of the attached line element and/or the coupling from external elements. The net reflected wave can be represented by an element reflection coefficient $\Gamma_L(\alpha, \beta)$, in which α and β are the relative phase shifts between rows and columns of the array lattice, respectively. In addition, each generator-phase-shifter combination may be mismatched to the transmission line and introduce a second reflection coefficient Γ_g.

Taking into account both types of reflections, it can be shown that the peak gain of a large scanning array when the beam is scanned to some angle θ_o from normal to the plane of the array is given by

$$G_{max} = \frac{4\pi}{\lambda^2} n d_x d_y \cos\theta_o \frac{\left(1 - |\Gamma_g|^2\right)\left(1 - |\Gamma_L|^2\right)}{\left|1 - \Gamma_g \Gamma_L\right|^2} \qquad (C\text{-}1)$$

where n is the number of elements. To optimize the gain of a phased array, let

$$\Gamma_g = \Gamma_L = 0 \qquad (C\text{-}2)$$

so that

$$G_{max} = \frac{4\pi}{\lambda^2} n d_x d_y \cos \theta_o \qquad (C-3)$$

The gain degradation due to random phase errors, caused by manufacturing tolerances and random mechanical distortion of the aperture surface, has been analyzed statistically by Ruze* and is discussed in Section 4.5.2. When the correlation interval, defined as that distance "on average" at which the phase errors become essentially independent, is large and where $e^{-\overline{\delta^2}}$, the mean square phase deviation in radians, is small, the gain degradation can be described as

$$\frac{G}{G_{max}} = e^{-\overline{\delta^2}} \qquad (C-4)$$

or

$$G = \frac{4\pi}{\lambda^2} n d_x d_y \cos \theta_o e^{-\overline{\delta^2}} \qquad (C-5)$$

where G is the antenna gain which accounts for aperture illumination phase perturbations. Thus, optimum gain is achievable when the mean square phase deviation is minimum and also by requiring that the radiator and generator-phase-shifter elements in the individual feed networks be matched as the beam is scanned. It also follows that the matched, individual generator-phase-shifters can be replaced by a matched corporate feed network to produce the same optimum gain response with matched radiators.

From Equation (C-5) it can be observed that the array gain will vary as the cosine of the scan angle, and for a very wide angle scan range of ±70 degrees, the maximum reduction in gain is 4.66 db. However, in general, the array aperture will not be matched for all

*See Reference 11, Section 4.6.

scan angles because $\Gamma_L \neq 0$ over this range. This mismatch can be taken into account by having $\Gamma_g = 0$ and Equation (C-1) then becomes

$$G_{max} = \frac{4\pi}{\lambda^2} n d_x d_y \cos\theta_o \left| 1 - |\Gamma_L|^2 \right| \qquad (C-6)$$

Similarly equation (C-5) then becomes

$$G = \frac{4\pi}{\lambda^2} n d_x d_y \cos\theta_o \left| 1 - |\Gamma_L|^2 \right| e^{-\delta^2} \qquad (C-7)$$

This equation accounts for the gain degradation effects due to the changes in the projected aperture, the impedance mismatch and the random phase errors.

APPENDIX D. HIGH GAIN, SELF-STEERING ANTENNA SYSTEM: ENGINEERING MODEL FOR SATELLITE-EARTH COMMUNICATIONS

An engineering model of a self-phasing antenna system for satellite-earth communications is being designed and fabricated at the Hughes Aircraft Company (NAS 5-10101). It is scheduled for test, evaluation, and delivery in 1967. This system, shown in block diagram in Figure D-1, incorporates two channels, each with a 125-MHz RF bandwidth, but for the sake of clarity, only one channel is illustrated completely in the figure. The system is designed to receive at 8 GHz and transmit at 7.3 GHz. The design is based on application to a gravity-gradient oriented and stabilized satellite in synchronous orbit with a conical coverage angle of ±15 degrees. This coverage allows

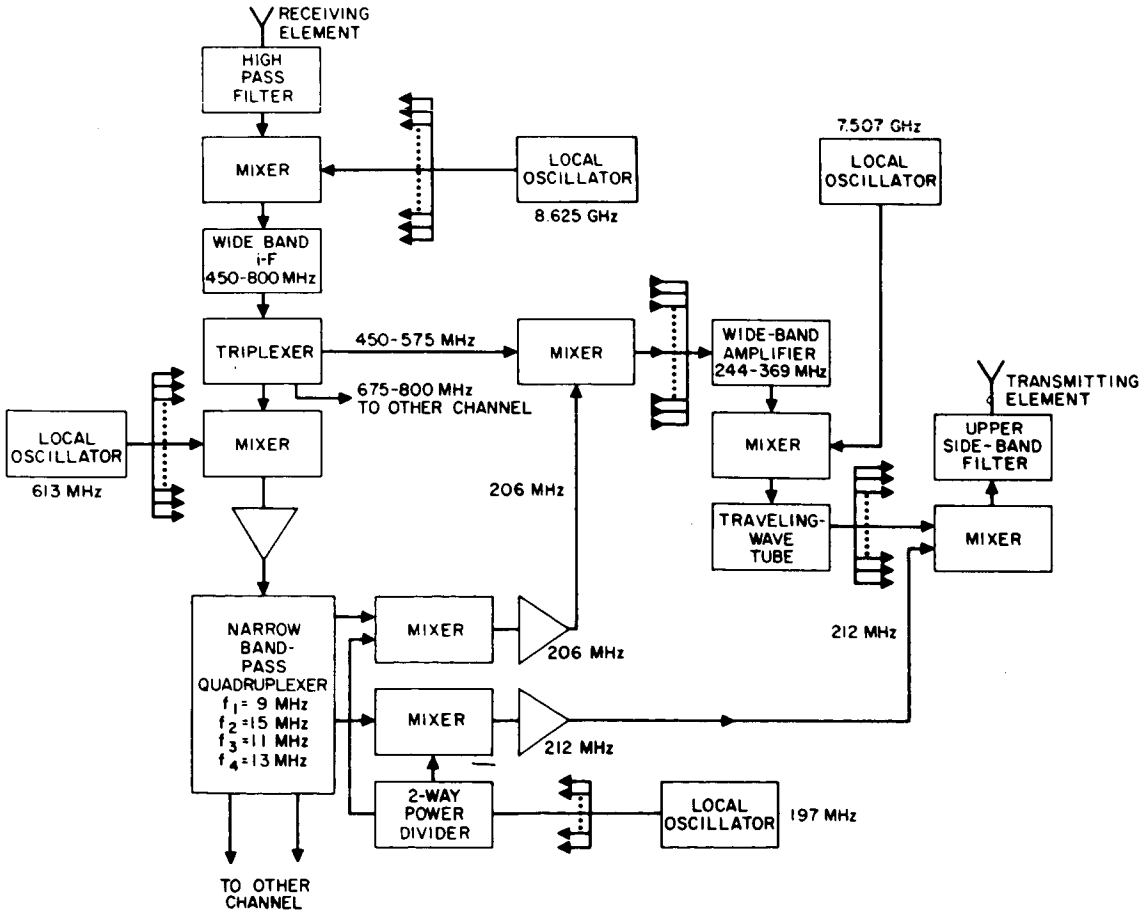


Figure D-1. High-gain, self-steering engineering model schematic.

for uncertainties in the attitude of the spacecraft. The transmitting and receiving portions will steer appropriate beams along arbitrary directions within that cone. Two independent channels will be provided, and four independent beams will be steered. The beam designations and the frequency bands utilized are shown in Figure D-2. Design objectives include a minimum gain of 30 db for the receiving mode and 25 dbw effective radiated power for the transmitting mode.

This system is intended to serve as a communication link to relay information transmitted from one station to another station via high-gain beams. The positions of these beams are controlled by the phase information obtained from CW pilot signals which are generated by the communicating ground stations.

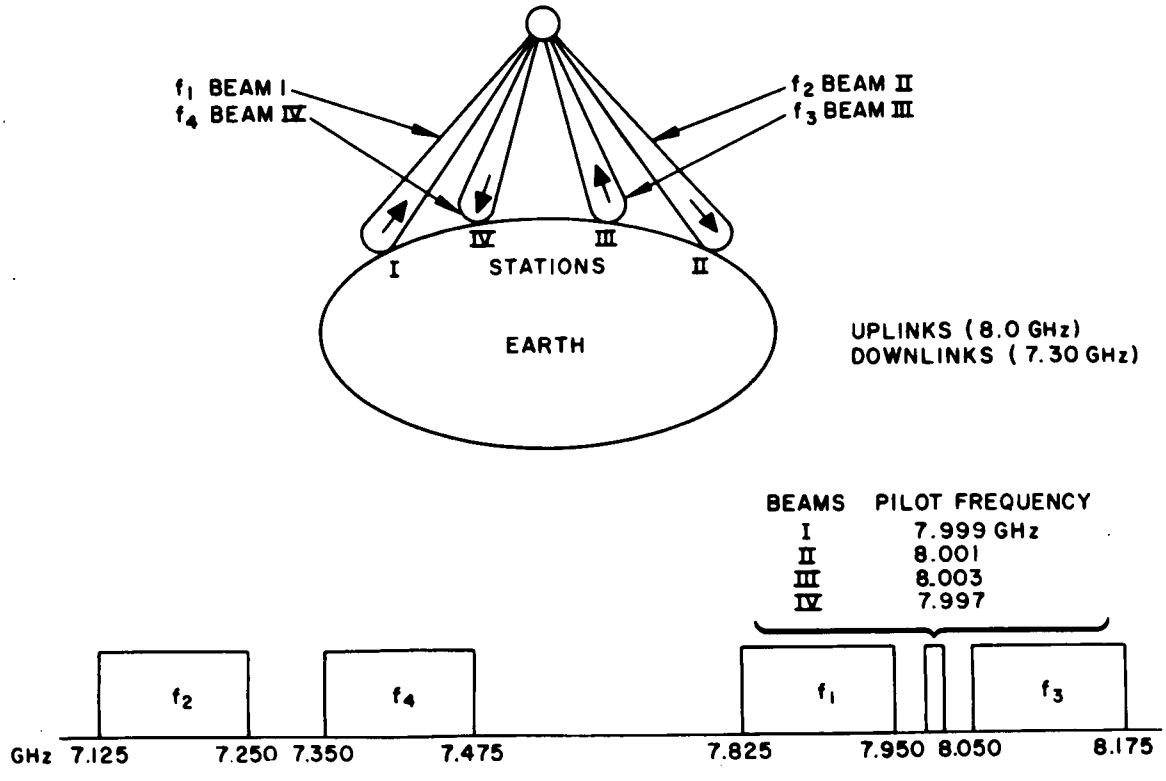


Figure D-2. Synchronous altitude gravity-gradient 30-degree cone of coverage.

For the channel of the engineering model shown in Figure D-1, a receiving pilot, a transmitting pilot, and a modulated signal are received by the receiving element, passed through a high-pass filter, down-converted to an intermediate frequency, and amplified by a wide-band IF preamplifier. After preamplification, the information signal and the pilots are separated by means of a triplexer filter. The pilots are then down-converted to a second, lower, IF to allow utilization of very narrow-band band-pass filters to establish a good signal-to-noise ratio for the pilots. These band-pass filters comprise the quadruplexer which, in addition to limiting the noise bandwidth of the pilot channels, serve to separate the pilot signals. After passing through the quadruplexer, the pilot signals are up-converted to about 200 MHz to enable these pilots to be mixed with the wide-band modulated signals without overlap of the power spectra.

With reference to Figure D-1, the modulated signal, 450 to 575 MHz, passes from the triplexer to a wide-band mixer, and the receiving pilot signal, 206 MHz, also passes to this mixer. The modulated signal is denoted by $\cos \{[\omega_c + f(t)]t - \phi_i\}$ and the pilot signal by $\cos [\omega_p t - \phi_i + \beta]$, where ω_c is the carrier frequency, ω_p is the pilot frequency, f(t) is a modulating signal, ϕ_i is the phase angle of the received signal relative to an arbitrary reference for the i^{th} element, and β is the phase shift of the pilot signal relative to the modulated signal, common for all elements. If these two signals are mixed and the lower sideband retained, there results $\cos \{[\omega_c - \omega_p + f(t)]t - \beta\}$; therefore, it is seen that the phase of the resultant IF signal is independent of the relative phase angle of the signals at the elements. The signals from the output of these mixers, (one for each element) which are in phase, are summed. At the point of summation, the receiver array gain is realized for the information signals.

The signal is then amplified at IF, up-converted to RF, amplified at RF, and then distributed to the final transmitting mixers. At these mixers the transmitting pilot is mixed with the modulated signal and the upper sideband is selected by the band-pass filter that follows. A modulated signal is produced at a transmitting element; this signal

has a phase angle which has the opposite sense from the phase angle of the transmitting pilot at the corresponding receiving element. The condition necessary to transmit the information from the antenna system in the direction of the transmitting pilot is that the recovery and transmitting arrays be scaled in wavelength.

Table D-1 presents the electrical and physical characteristics of the engineering model, and the projected characteristics for a flight model of a similar system. Figure D-3 shows the configuration for the engineering model. The flight model will be configured similarly.

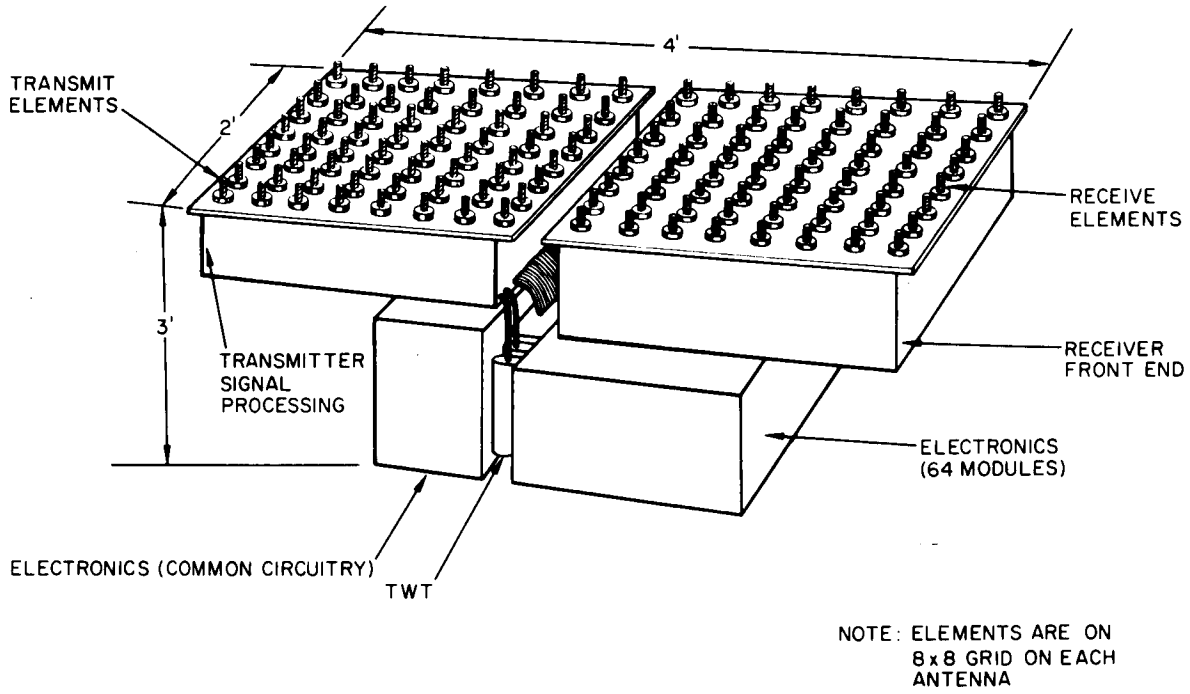


Figure D-3. Engineering model in artist's conception.

	Engineering Model	Projected Flight Model
Number of elements (each for 2 arrays)	64	64
Number of channels	2	2
R-f bandwidth (each channel)	125 MHz	125 MHz
Guard band (between channels)	100 MHz	100 MHz
Up-link frequencies (center) Channel A	7.8875 GHz	7.8875 GHz
Up-link frequencies (center) Channel B	8.1125 GHz	8.1125 GHz
Down-link frequencies (center) Channel A	7.1875 GHz	7.1875 GHz
Down-link frequencies (center) Channel B	7.4125 GHz	7.4125 GHz
Total cone angle of coverage	30°	30°
Element gain (minimum)	11.6 db	11.6 db
Array gain (minimum)	29.8 db	29.8 db
Polarization	Circular	Circular
Effective radiated power (objective)	25.0* dbw	33.0 dbw
Ratio of pilot to modulated signal power when 125 MHz bandwidth is utilized	-13.0 db	-13.0 db
Received modulated signal level required for 19-db minimum SNR of transmitted signal when total bandwidth is utilized	-115.5 dbw	-115.5 dbw
Power consumption: receiver	32.0 watts (excluding local oscillator)	21.5 watts (including local oscillator)
Power consumption: transmitter	201.1 watts (excluding local oscillator)	204.5 watts (including local oscillator)
Power consumption: Pilot processor	108.7 watts	108.7 watts
Power consumption: Attitude readout	0.9 watt	0.9 watts
Power consumption: Power supplies	--	73.3 watts
Power consumption: Total prime power	--	408.9 watts
Total weight	175 pounds (excluding microwave local oscillators and all power supplies except the traveling-wave tube power supplies)	175 pounds (including all power supplies and local oscillators)

*It is expected that 33 dbw will be obtained with the engineering model.

Table D-1. Characteristics of self-phasing array - engineering model.

APPENDIX E

TIME DELAY COMPENSATION FOR VERY LARGE ANTENNA ARRAYS

INTRODUCTION

Some of the special problems of large ground arrays are discussed in this appendix. The array is considered somewhat independently from the remainder of a communication system in which it is a part. That is, it is not the point of this appendix to prove the need for a large ground antenna because this is the only way of achieving high bit rate data but rather to assume a large antenna is needed and to discuss it in the context of a high data rate system. Such an approach allows flexibility in the examination of an antenna (or array of antennas) which is considerably larger than any which are currently under construction. In this approach, and in particular the discussion which follows, several choices are made initially which are somewhat arbitrary and lead to certain limitations. However the questions discussed and analyzed are certainly those which must be considered when large antenna systems are constructed.

Initial System Design Parameters

The large antenna to be considered is taken to be square and 2000 feet on a side. The bit rate assumed is 10^6 bits per second. Assuming such a bit rate allows constraints in time delays, etc., to be set but also leads to performance limitations at very low bit rates.

A radio frequency of 2300 MHz is assumed.

The array is taken to be used primarily for data gathering with functions such as command transmission and angle tracking relegated to a secondary antenna system which can be of conventional size. (Additional investigation of methods for angle tracking is required as noted in Section 7.2.)

This large antenna system will have many signal processing features, for these and for other reasons discussed below a computer (probably a general purpose computer) is assumed to be available. The uses for this computer will be discussed in the appropriate sections.

Antenna System Design Problems Considered

The primary problems considered in this Appendix are: 1) the problem of receiving, in phase, the energy impinging on the antenna. 2) the problem of compensating for the difference in time of arrival of the wavefront to different points on the antenna array, and 3) the expected relative performance of two methods of processing data from the antennas.

In addition to the problems listed above, several design choices are derived as a result of the basic assumptions such as the size of these apertures and required degree of time correlation.

PRELIMINARY SYSTEM DESIGN

The major design choices of the large aperture antenna are documented in this section. These are a consequence of the size, bit rate and radio frequency choices already noted.

Subaperture Size

The problem of correcting the RF phases of signals received from a number of distinct subapertures prior to summation is essentially a classical antenna problem. By contrast, the correction of the differential time delays constitutes an essentially distinct problem which can, perhaps, best be considered as a communications system problem rather than an antenna problem. It will become apparent that some of the most attractive means of compensating for the differential time delays in a very large antenna array will simultaneously provide an attractive means of dealing with the need for appropriate RF phasing prior to signal summation. The resulting phasing techniques, while by no means the only techniques that could be employed, may prove to be the most

attractive means for dealing with the phasing problem. Here a case is encountered in which solution of a system problem necessitates use of a technique which simultaneously solves an entirely different problem in one of the elements of the system (in this case, the antenna).

At this point it is appropriate to consider briefly the major options available as to choice of subapertures. The choice is very wide and ranges, in principle, all the way from individual dipole elements to parabolic dishes with diameters of 200 feet, or possibly more. In fact, these two examples constitute extreme cases, both of which are impractical for different reasons. The sheer number of individual dipoles required for a very large total aperture is so great that any attempt to treat each individual dipole as a separate receiving subsystem or subaperture with its own provision for time delay compensation becomes prohibitively complex and expensive. In addition, at the frequencies of interest, even provision of a separate mixer and transmission line for each dipole is unattractive by virtue of the very large numbers of these elements required. Finally, the receiving cross section of an individual dipole element will be too small to permit reception of a reasonable amount of signal power; the signal-to-noise ratio with which any signal processing equipment associated with the individual aperture would have to function would be so small as to require an impractical dynamic range of operation. In the same vein, the signal-to-noise ratio would be far too small to consider, for example, use of a separate phase-locked receiver at each such subaperture. The significance of some of these comments will become clearer as the design of a possible array is described in greater detail.

The use of parabolic dishes 200 feet or more in diameter is also unattractive by virtue of cost. Here is a situation in which the cost of the aperture itself and the pointing and drive mechanism associated therewith becomes exorbitantly large relative to the signal processing equipment associated with the antenna.

From the preceding discussion, it is apparent that an intermediate subaperture is required. The obvious possibilities for the individual subapertures are conventional gimbaled dishes of a suitable intermediate size or some form of dipole or planar array. While conventional gimbaled structures could be employed, the large numbers required for very large total receiving apertures again implies an excessive cost when account is taken of the costs associated with the gimbal and drive mechanisms. This consideration indicates the desirability of employing an ungimbaled subaperture. Such a subaperture will then be mounted in an essentially horizontal position and, since it should be large relative to the wavelength, will require some form of electronic scanning. So long as the size of the subaperture is not chosen to be excessively large, the beamwidth will not be so small that steering the beam of the individual subaperture requires holding excessive tolerances. The beam steering will be accomplished by computer control.

The decision that the preferred subaperture will be ungimbaled implies that some loss in effective receiving cross section will occur for sources that are not located in the neighborhood of the zenith. For a well designed subaperture, this reduction in aperture efficiency will vary as the sine of the elevation angle. Thus, for a source 30 degrees above the horizon the aperture efficiency will be approximately 3 db lower than for the case when the source is at zenith. In spite of this loss, the use of ungimbaled apertures should be sufficiently less expensive that overall cost consideration should be expected to favor this configuration. Another point is that horizontal subapertures can be closely clustered without interfering one with another, whereas large gimbaled structures, if closely clustered, will block one another's field of view when tracking sources which are not close to zenith. Thus, gimbaled apertures require at least as extended an array as do horizontal subapertures and are subject to an increased grating lobe problem (see Section 4.5).

System Influences on Subaperture Size. The fact that subapertures are ungimbaled and horizontally disposed in itself immediately imposes a significant constraint. To appreciate this constraint, it will be convenient to suppose that the information received from the spacecraft is coded in a digital format. This is not an essential assumption inasmuch as an argument similar, but not identical, to that to follow can be made for the case of any choice of modulation form. The fact that digital data transmission will be employed to a very large extent on future space missions makes the digital case preferred for illustrative purposes.

To be useful, a very large antenna array should be usable at least down to elevation angles as low as 30 degrees* relative to the horizon. At such low elevation angles, the signal received at the rear edge of the aperture will lag that received at the front edge of the aperture by a time approximately equal to the time required for an electromagnetic wave to propagate a distance equal to the aperture dimension. The fact that the RF phases of the signals received at various points on the aperture will differ will be corrected by the means employed for electronic scanning. However, unless the aperture itself is designed to incorporate pure time delay phase compensation, the corresponding intelligence received at the far edge of the aperture will still lag that received at the forward edge of the aperture by approximately the time required for an electromagnetic wave to propagate across the aperture dimension. When this propagation time is significant relative to the bit length, the modulation received at the subaperture output will be distorted in the same manner as that described earlier for the case of signals received at individual subapertures. Since pure time delay compensation within a subaperture is almost certainly not feasible on the basis of cost and complexity, it follows that the maximum dimension

*Elevation angles as low as 15^{o} are useful but would be quite difficult for a phased antenna to accomplish without serious sidelobe problems.

usable for a subaperture must be small relative to the equivalent spatial length of one bit period; that is, the distance travelled by an electromagnetic wave during the bit duration. Detailed analysis will be required to determine the extent of garble and modulation distortion introduced as a function of aperture dimension. A reasonable choice of aperture dimension, would be such that the aperture dimension did not exceed about one tenth of the spatial equivalent of a bit period at the highest bit rate for which the receiving system is to be designed. Beyond this point the system's self noise should become noticeable, while with aperture transit times not exceeding one tenth of the bit period the integrated effect over the aperture should be essentially negligible.

The preceding consideration sets a maximum dimension for a subaperture. Depending upon the particular choice of the maximum bit rate for which the aperture is to be designed, this subaperture dimension may or may not exceed the largest practical subaperture dimension.

The spatial equivalent of one bit at a one megabit per second data rate is very nearly 1000 feet. Employing a rule of thumb that the dimension of the subaperture should not exceed approximately one tenth of the spatial equivalent of a bit period indicates that the individual subarrays should not be more than about 100 feet on a side. Any method of combining elements or sub-subarrays which will provide suitable electronic scanning will prove satisfactory so long as this dimension is not exceeded. Moreover, a dimension of this order seems a very reasonable basis for design since the tolerance requirements for a subarray of this size can be met relatively easily at DSIF frequencies. This fact can be appreciated by observing the number of 85-foot paraboloids which have been built to function at the DSIF frequencies and above, and by noting that the tolerances for the planar structure will provide about a factor of two latitude relative to those for the paraboloid. In practice, it may prove convenient to fabricate each 100-foot subaperture of sub-subapertures

of perhaps 25 or 50 feet on a side. The decision implied here should be made by the supplier and need not be cause for further concern at this point.

In summary, a 100 foot subaperture constitutes a very reasonable basis for design, and a subaperture of this size will be supposed hereafter. This subaperture is certainly sufficiently small that decorrelation caused by wavefront distortion due to tropospheric and ionospheric inhomogeneities will not be a significant factor; this fact is adequately established by extensive experience with large dishes at these frequencies. Even with subapertures as large as 100 feet on a side, a total of 400 subapertures would be required for a receiving system whose total array area is equivalent to that of a single array 2000 feet on a side. This fact illustrates dramatically the need for subaperture designs which minimize cost and thus shows the economic infeasibility of clustering, say, 100-foot paraboloids to achieve an equivalent receiving area.

Subaperture Sidelobe Control. Still another factor to be considered in subaperture design is the question as to whether or not precise sidelobe control is required for the individual subaperture. In general, it is to be expected that a suitable illumination taper will be employed across the full composite aperture for sidelobe control. However, two alternatives present themselves. The first alternative is to provide a suitable illumination taper across the subaperture, sufficient to provide the desired sidelobe characteristics. The remaining possibility is to employ a combination of subaperture illumination taper and illumination taper across the composite aperture. While the question as to which of these techniques should be employed constitutes an important matter for further investigation, it will not be dwelt upon in this section because it is a matter of concern primarily for the design of the full antenna array rather than peculiar to the problem of time delay compensation across such a large array. As will be seen shortly, however, the two questions

are not completely independent inasmuch as with some configurations rather good control over the sidelobes of the individual subapertures may prove necessary.

Phasing Subapertures

The heart of the carrier phasing problem lies in the means employed for initial acquisition. This is the case because, once a suitable composite carrier reference is available, the carriers received from each individual subaperture can be compared against the composite reference carrier and their phases adjusted accordingly. Thus, once acquisition has been effected, any of a number of conventional closed-loop control techniques can be employed to assure that the carriers are thereafter appropriately phased relative to one another. In view of this fact, the means which can be employed for initial carrier acquisition is considered now.

Three methods of phasing the individual subapertures into a composite array are discussed. They are: 1) the use of phase locked receivers at each subarray 2) the use of statistical combinations of the outputs of the arrays to obtain a reference signal and 3) the use of a priori knowledge of the signal source to set the proper phase in each subaperture.

Phase Locked Loop Phasing. The most straightforward means of acquiring the carrier is to employ a separate low noise phase-locked receiver for each individual subaperture. Acquisition by this technique implies that sufficient carrier power must be available in the output from the subaperture to permit lock-on and this requirement, in turn, may under some circumstances have significant implications for the division of signal power between the carrier and the information modulated onto it. Clearly, any signal power which must remain in the carrier to facilitate initial acquisition is no longer available for information transmission. In the extreme case situations are conceivable in which the signal strength at the output of the subaperture would be insufficient to permit carrier lock-on even if the full power

of the signal resided in the carrier. This might well be the case for a space vehicle operating at such an extreme range that information rates of only a few bits per second would be achievable even with the full aperture of the composite array. It is also important to observe that, when a separate phase-locked receiver is employed at each subaperture for carrier acquisition, careful attention to sidelobe control may be required for the individual subaperture. That is, amplitude taper over the complete array may have very beneficial effects for the array as a whole, but will do nothing to control the sidelobes, and hence the received backlobe or sidelobe radiation in the individual receivers. Since the lock-on characteristics of the phase-locked receiver associated with the subaperture are influenced only by the signal to spectral noise density in the subaperture output, sidelobe control to assure a low level of backlobe and sidelobe thermal radiation may well prove essential.

It is important to determine how much of a restriction the use of a separate phase-locked receiver associated with the individual subaperture imposes on the acquisition capability of the receiving system. By means of an elementary argument it is possible to show that for the high data rates of interest, carrier acquisition using the signal received at each individual subaperture poses essentially no restriction. In practice, a single phase-locked receiver located at just one of the subapertures would suffice, with a coherent local oscillator or carrier reference signal relayed to the other subapertures for data frequency translation or demodulation. Since the cost of a phase-locked receiver should be very small relative to that of the individual subaperture and since the use of a separate phase-locked receiver for each subaperture eliminates a number of other potential problems this design has been chosen for present purposes.

Specifically suppose that information is transmitted in a PCM format and that the PCM data are modulated onto the carrier using biphase modulation. The modulation form assumed is residual-carrier biphase. Hence the phase difference between a "0" and a "1" is

somewhat less than a full 180 degrees with the consequence that a small portion of the signal power remains in the carrier. For example, if instead of modulating the carrier by ±90 degrees, the carrier phase is modulated by ±70 degrees, the residual power remaining in the carrier will be about 10 percent of the total signal power. Leaving 10% of the total signal power in the carrier corresponds to a loss of power in the modulation (information) of about 0.5 db, which is certainly small. Modulating with phases even closer to 90 degrees (positive and negative) will, of course, leave even less power in the carrier.

Consider now how much power must be left in the carrier for successful acquisition. To place this question in proper perspective, observe that there is little point in acquiring the carrier if the signal-to-noise ratio at the output of the complete composite aperture is so low as to result in an excessive bit error rate on the PCM data. Thus, the signal-to-noise ratio at the output of the composite receiving system is sufficient to assure an acceptably low bit error rate. A bit error probability of 0.01, which corresponds to an output signal-to-noise ratio of about 7.4 db, is probably excessive for any application in which a great deal of data is to be transmitted. A bit error probability of 0.001, which corresponds to an output signal-to-noise ratio of about 10 db, is about the worst bit error probability one would expect to tolerate for such high bit-rate data. At signal-to-noise ratios much larger than 10 db, the bit error probability becomes almost vanishingly small. Thus, the output signal-to-noise ratio of interest for PCM data reception for the complete receiving system must be in excess of 7 db, with a value of 9 db as probably a more realistic lower acceptable limit. This output signal-to-noise ratio is to be measured in a bandwidth equal to one-half the bit rate in a matched (integrate and dump) detector.*

*Analysis shows that the noise bandwidth of the complete receiving system should be taken as one-half the bit rate, in the data channel, when optimum demodulation, employing an integrate-and-dump circuit, is used.

Next consider the corresponding signal-to-noise ratio for the demodulated data at the output of a phase-locked receiver associated with a single subaperture. On taking the foregoing figure of 9 db as representing the threshold condition, it is seen that the signal-to-noise ratio of the data at the output of the subaperture receiver will be reduced by a factor of 400, or 26 db, since the composite array contains some 400 subapertures. The resulting signal-to-noise ratio in a 500 kHz (half the bit rate) bandwidth will thus be 9-26 = - 7 db. Consider now the signal-to-noise ratio that would be obtained in the carrier phase-locked loop if the signal power were equally divided between the carrier and the modulation. The bandwidth of the carrier phase-tracking loop should not exceed 10 Hz for a well-designed receiver. Since the ratio between the output noise bandwidth of the digital data (500 kHz) and this 10 Hz carrier loop bandwidth is 50,000, or 47 db, it is seen that the -17 db signal-to-noise ratio at the output of the phase-locked receiver in the information bandwidth corresponds to an output signal-to-noise ratio from the phase-locked loop of at least -3 - 17 + 47 = 27 db. This signal-to-noise ratio in the phase-locked loop bandwidth is far more than adequate for solid lock. In fact, a 6 db signal-to-noise ratio in the loop noise bandwidth will provide solid lock, while a 10 db signal-to-noise ratio will provide a relatively clean reference carrier.

From these considerations, it follows that splitting the power equally between the carrier and the modulation produces a signal-to-noise ratio in the carrier phase-tracking more than 20 db larger than is required. This fact, in turn, implies that the percentage of the power in the carrier can and should be reduced. If the carrier power is reduced by another 10 db, the loss in signal-to-noise ratio in the modulation output will be negligible relative to a pure biphase system, in which the carrier is completely suppressed. At the same time, sufficient excess power will remain in the carrier so that acquisition

can be effected by a phase-locked receiver associated with a single subaperture even for the case where the data rate is reduced to 100 kilobits per second, with threshold operation implied for the resultant bit error rate at the output of the composite receiving system. For that matter, by splitting the power half and half between the carrier and the modulation, very solid carrier lock can be obtained at an individual subaperture even for the case where the data rate is reduced to 10 kilobits per second, with near-threshold bit error rates at the output of the composite receiving system. These considerations illustrate the fact that carrier acquisition by a phase-locked receiver associated with each individual subaperture will pose no problems or limitations for carrier tracking even when the composite receiving system is operated at only a 10 kilobit per second data rate with threshold signal-to-noise ratios.

The preceding argument has made no attempt at great precision. A more refined argument of the sort given above could be employed to determine the lowest data rate at which the composite system could be operated while still employing the signal received by only a single subaperture for carrier acquisition. This refinement is unnecessary at this point as our interest is primarily in high data rate deep space communications, and low data rate deep space communications with such a receiving system, which would correspond to exploration of the outer reaches of the solar system, should be regarded as a fallout capability at the present time.

Phasing Using Statistical Combinations. A recent investigation has shown that, if each signal from a subaperture is split into two signals whose phases differ by 180 degrees and if, for a total of n subapertures, the 2^n sums formed by choosing one of the two signals from each subaperature and adding, in all possible ways, are formed, at least one of the resulting 2^n signals will exhibit a carrier component at least equivalent to that which would be obtained with a

composite aperture 40 percent as large as the composite formed from the total of n subapertures, supposing perfect phasing*. For large numbers of subapertures, say on the order of several hundred, the resulting number of combinations clearly becomes so large as to render this method of acquisition infeasible if an equivalent 40 percent efficiency is to be achieved.

An alternative would consist of introducing phase shifts chosen from a table of random numbers in the outputs of each of the subapertures and adding the resulting signals. By repeating this process only a modest number of times, it seems likely that a carrier component sufficient for acquisition could be achieved with relatively high probability. This possibility requires further investigation. The analysis should prove entirely straightforward inasmuch as the random phase shifts introduced in each summand prior to summation convert the problem into the classical problem of random flights (Rayleigh), which is perhaps better known in contemporary circles as the random walk problem. It is readily seen from this observation that the resulting amplitude of the carrier component will have a Rayleigh distribution.

It is, perhaps, important to observe that the same technique can be employed to any collection of the subapertures, rather than to the totality thereof. By this means it may in some cases be feasible to assure carrier lock-on without the complications inherent in handling the signals from all the subapertures simultaneously for this purpose. In this connection, the phase-splitting technique described previously might prove feasible if a relatively small number of subapertures were employed. It should also be observed that either the phase-splitting or the random phase shift technique could be employed in time sequence

---

*Half of the resulting 2^n sum signals are redundant since to each of the sum signals there corresponds another differing only in sign. It follows from this observation that, in practice, only a subset of 2^{n-1} signals would be formed. For further information see Reference 1, page 250, and Reference 2.

rather than simultaneously in parallel channels if the time available for acquisition permitted. Sequential acquisition by this means would reduce the amount of equipment required and hence would result in very substantial cost savings. In any event, some care should be exercised to avoid an acquisition scheme utilizing more subapertures than are really required to provide an adequately clean carrier reference signal since unnecessary proliferation can prove both complex and costly.

Phasing Using A Priori Information. Acquisition by summing the outputs of a number of the subapertures to obtain a reinforced carrier component can, up to a certain point, be performed on the basis of a priori knowledge. That is, up until the total dimension of the region scanned by a collection of subapertures becomes so large that decorrelation caused by atmospheric or ionospheric inhomogeneities become significant, the phase corrections to be introduced in the outputs of the individual subapertures can be determined on an a priori basis from a knowledge of the angular position of the source. The resulting phase corrections can then be made by means of variable phase shifters prior to addition. As a practical matter, the number of subapertures which are ganged in this fashion will be limited by the phase stability of the various circuits and transmission lines and by the accuracy with which the angular position of the source is known; the phasing problem for a collection of subapertures is equivalent to the beam-pointing problem for a single antenna whose dimension is equal to the maximum separation between the subapertures in the group employed in this way.

Time Delay Correlation Determination

Since a phase-locked receiver will be provided at each subaperture, only video data need be transmitted from each subaperture to a central point for time delay compensation and correlation to enhance the system output signal-to-noise ratio. Whatever technique is employed for data transmission, it will be necessary to provide

some means for assuring adequate knowledge of the relative timing of the various signals collected at the central station. There appear to be two possibilities for obtaining this information. First, a priori information can be employed. Secondly, relative timing data can be obtained directly from the received signal. These two possibilities will be discussed briefly in the next paragraphs.

A Priori Time Delay Correlation Determination. It should be observed that the available timing tolerance is of the order of 0.1 microseconds for the highest data rate (one megabit per second) for which the receiving system is designed. Since the transmission lines employed will be of the order of 2000 feet in length, maximum, it follows that uncertainties in the delays introduced by these lines should be at most a small fraction of this tolerance. In a similar vein, the variations in time delay associated with both active and passive circuit elements can be kept very small by careful design. A variation of at most a few nanoseconds could be achieved. The only other source of differential delay requiring consideration is the differing times at which the received signal reaches the various subapertures. However, since the relative geometry of the system will be known to a small fraction of a foot and since the angular position of the source, when tracking a deep space vehicle, will be known to a small fraction of a milliradian, it is apparent that no timing errors which are significantly large relative to the 0.1 microsecond tolerance can arise due to this cause. It thus appears that the relative timing of the signals available at the outputs of the various transmission lines (400 in number) associated with the individual subapertures will be known with more than adequate accuracy to permit time delay compensation to the requisite tolerances prior to summation. The shortcoming of the a priori method of time delay compensation is it does not provide a

means of removing variable time delays such as could be due to atmospheric variations nor does it provide a means of removing slow variations in "fixed" delays such as would be due to the transmission lines. Clearly for optimum performance a means must be provided for removing the variable time delay. This is discussed in the section which follows.

Signal Information for Time Determination. An alternative to the use of a priori data, as discussed above, is to derive the differential timing reference information from the received signal itself. This can be done relatively simply in either of two ways. The first way consists of transmitting an auxiliary timing waveform, suitably coded, from the spacecraft and extracting this waveform on the ground with a suitable narrowband synchronizing system. For example, the timing information could be imposed on the transmitted signal in the form of the phases of one or more subcarriers. After demodulation of the received signal to obtain the baseband structure, these individual subcarriers can be phase-tracked in very narrowband circuits to recover the required phase (and hence time) information. The alternative, and esthetically somewhat more satisfactory, method is to take advantage of the fact that timing information is already provided in the PCM data transmission. That is, some means is usually provided in a PCM system for achieving bit, word, and frame synchronization. By taking account of the known characteristics of the synchronizing information built into the PCM frame structure, very narrowband systems can be employed to obtain synchronization on the output of each of the individual receiving systems. These synchronizing systems must, of necessity, be extremely narrowband because of the low signal-to-noise ratios in the PCM data outputs of the individual receivers. This point was discussed previously, where it was shown that the signal-to-noise ratio in the output PCM data from a single receiver might be of the order of -20 db when the overall system is near its threshold. The fact that the signal-to-noise ratio in the individual PCM outputs is this low implies the need for a relatively long correlation time in relation

to the bit bit rate to achieve adequate synchronization. This long correlation time may prove to be a distinct disadvantage. This fact, in turn, renders transmission of a separate timing waveform, in the form of, perhaps, sinusoidal subcarriers, more attractive. Once timing information is provided in this form, however, it is no longer necessary to provide the PCM data with built-in synchronizing signals in the form of bit, word, or frame synchronizing patterns.

Clearly, the advantages and disadvantages for the derivation of timing information using a priori information or signal derived information require careful study and comparison to determine the best. The a priori method seems at this time to be simplest and adequate at the present writing and is assumed to be the method used. The short discussion of the alternatives has been included to show that, even if the a priori information were of insufficient accuracy, the required timing information could be obtained from the received signal itself. It is worthy of note that, even if special timing subcarriers were to be modulated onto the signal prior to transmission from the spacecraft, relatively little power would have to be placed in these subcarriers inasmuch as their known structure lends itself readily to phase measurement with extremely narrowband processing techniques. Additional equipment required for obtaining the requisite timing information from the received signal would significantly increase the cost and complexity of the composite receiving system. For this reason, a preferred system design is to function on the basis of the a priori differential timing data, which it appears can be obtained so readily.

Signal Processing the Subarray Signals

Basic Requirements. This section is devoted to a discussion of processing signals received from the various subapertures in such a way as to compensate the differential delays and enhance the resulting signal-to-noise ratio at the output of the composite receiving system.

In its most obvious form, this signal processing consists merely of the introduction of suitable time delays at the outputs of the individual transmission lines from the phase-locked receivers associated with the individual subapertures and summation of the resulting time-compensated signals. In addition to the compensation, provision must be made for maintaining equal signal levels at the output of each transmission line; that is, a suitable AGC must be provided. The time delay compensation at the output of each transmission line will, in general, consist of a fixed delay whose function is to equalize the time delays for a source at some fixed position, preferably at zenith, and a variable time delay mechanism. In the case where a priori data is used for time delay compensation, the variable component of time delay is controlled by the central control computer in such a fashion as to account for path length differences resulting from departure by the central control computer in such a fashion as to account for path length differences resulting from departure of the angular position of the source from the reference position for which the lines are equalized. In the case where the timing data are obtained from the received signal itself, the variable time delays are controlled in closed loop fashion by comparing the relative epoch at the output of each of the transmission lines with that at the output of a fixed transmission line whose output is chosen as a timing reference for the entire system. The timing errors are then corrected by a conventional feedback technique. As indicated previously, this latter means of compensating the time delays presently appears more complex. In any event, the signal-to-noise ratio after summation of the n ($n = 400$) time-compensated signals from the various transmission lines will be improved by a factor of n which, in this case, implies a 26 db improvement.

Methods for Achieving Variable Time Delays. In order to time correlate the signals from the 400 subapertures a suitable variable time delay must be available. This delay may operate as an analog time delay or as a digital time delay in that the circuitry required to

convert from one form to another would be a relatively small part of the overall antenna system cost. Described below are several possible time delay implementations including lumped constant delays, analog delays, dispersion delays, and digital shift register delays.

Lumped Constant Time Delay. The maximum linear dimension over which the various subapertures are dispersed will be a few thousand feet. Even in cases where the subapertures are not abutting, the maximum delay variation necessary will be of the order of a very few microseconds. For example, in the case where the subapertures abut, so that the maximum dimension of the array is along the diagonal of a square some 2000 feet on a side, the delay variation required will not exceed ±3 microseconds, or a total variation of 6 microseconds. In this connection it is important to note that only the variation is significant and that absolute delays much larger can be tolerated. A straightforward, although not very elegant, means of implementing the required variable delay would be by means of lumped constant delay lines of various lengths, all of which were binary multiples of a suitable fraction of the allowed 0.1 microsecond delay compensation tolerance. These lines could be switched into the desired configuration by a simple computer-controlled switching matrix.

The lumped constant time delay method of implementing the variable delay line would constitute rather a brute force solution. It might also prove somewhat undesirable because of the transients associated with reconfiguring the lines to vary the delay.

Analog Delay Techniques. For delays as small as ±3 microseconds various analog delay techniques are directly applicable. For example, a quartz acoustic delay line has been constructed the length of whose delay path is mechanically variable. If purely electronic control of the delay is desired, a ferrite line, in which the delay is varied by variation of a magnetic field, can be employed. Should delays

as large as 6 microseconds prove difficult to achieve with a continuously variable line, a tapped discrete line with taps at relatively large increments, say of the order of one microsecond, could be employed with a continuously variable line having a total variation of one microsecond in series. With this configuration, the switching transients would occur so infrequently as to be of very little concern.

Dispersive Time Delays. Dispersive delay lines have been built for use in pulse compression radars. By choosing the delay-versus-frequency slope of such lines to be sufficiently small and by modulating the desired intelligence onto a variable frequency carrier, it is possible to use a dispersive line as a continuously variable delay device. If carriers in the UHF region are not objectionable, YIG (yttrium iron garnet) lines can be used for this purpose. A more attractive kind of line, which will operate with a carrier frequency in the intermediate frequency range, is the tapered metal strip delay line such as has received considerable development attention at Hughes Fullerton.

Digital Shift Register Time Delays. The preceding discussion shows that the variable delays required for data time compensation can be effected with devices which are well within the state of the current circuit art. There is one other class of delay device which looks so promising for this application that it deserves special emphasis. The device in question is the digital shift register. Such a shift register can be used as a variable delay device by the expedient of indexing a bit stream through the shift register with a varying clock frequency to provide the requisite delay variation. A delay line of this sort is extremely attractive for the application contemplated because it is simple, reliable, and has a delay which is readily controlled to almost any required accuracy.

Consider the requirements on a shift register delay line capable of accepting a one megabit per second input and providing a controllable delay variation over a range of six microseconds. If the data from the individual phase-locked receivers is coded such that each binit is approximately one bit of information, a single digital shift register delay line will provide the required delay compensation at the output of each receiver channel. If the data are transmitted to the central station in analog form and then sampled and converted to digital form in such a fashion as to preserve the amplitude distribution of the receiver output (that is, if no attempt is made to make a zero-or-one PCM bit decision at the individual receiver outputs), it will be necessary either to use a separate line for each bit in the PCM word or to operate the line at a clock rate higher by a factor equal to the number of bits in the word, so that the entire bit stream can be passed by a single shift register.

One point concerning the use of shift registers with variable clock frequencies for variable delay lines warrants special notice. The point in question is that the clock rate of the variable digital delay line is nonsynchronous with respect to the incoming bit stream, whether this bit stream be the bit-for-bit estimate of the PCM output of the phase-locked receiver or a sampled and quantized version of the analog output of the receiver. This must necessarily be the case since the clock rate must be variable in order to provide a controllable delay. This fact implies, in turn, that the digits in the input bit stream will be sampled at varying positions within a bit, with the consequence that the output of the digital delay line will not preserve precisely the beginning and end of any particular bit. This, in turn, means that a certain amount of noise will be introduced into the output signal prior to combining it additively with the other delay-compensated signals. Such noise will have the effect of degrading the operation of the composite system. One way of circumventing this problem would be to employ a clock frequency high relative to the incoming bit rate, so that the leading and trailing edges of the individual bits would be

preserved to an adequate degree of accuracy. Using the same rule of thumb employed in selecting the maximum possible subaperture size, we might require a minimum clock rate for the delay line at least 10 times the bit rate of the incoming digital data. In view of the possibility of a tolerance buildup problem among all the sources of a delay error, a more likely requirement would be a clock frequency at least 20 times the incoming bit rate. Thus, for a one megabit per second input to the line, the minimum clock frequency required might be of the order of 20 MHz.

Suppose, for illustrative purposes, that the line is designed for a minimum clock frequency of 20 MHz proposed above. A reasonable maximum bit rate for the line must be determined before sizing the line to determine how many stages should be provided in the shift register. A reasonable choice would seem to be to fix the upper clock frequency at twice the minimum clock frequency; that is, to set the upper clock rate at 40 MHz. With this choice of upper and lower clock frequencies, the delay effected by the line will vary by a factor of two. Since the line is to provide a total bit variation of six microseconds, the maximum absolute delay which it must provide will be 12 microseconds, and this delay will be realized at the 20 MHz clock frequeny. It follows from this observation that the shift register must contain a total of 120 stages. This is not a particularly large shift register by current standards, especially in view of the fact that 100-stage shift registers are presently available on a single silicon chip mounted in a TO-5 can. While it is true that these large shift registers on a single chip are presently MOSFET (metal oxide silicon field effect transistor) devices which will not presently operate at clock frequencies higher than about two megabits per second, the existence of these devices provides a basis for optimism that a device with the desired characteristics should be available on a single silicon chip within a few years at a modest price. In any event, such shift registers can presently be constructed either with integrated (monolithic) circuits or with discrete semiconductor components.

been preserved in analog* form up to and including summation to enhance the system output signal-to-noise ratio, and that the resulting PCM data were recovered following summation. An interesting alternative is to digitize the output of the phase-locked receiver associated with each individual subaperture prior to "summation" to obtain an enhanced signal-to-noise ratio. The procedure in this case would be as follows. The receiver output in the modulation band would be passed through a decision circuit and converted into a train of zeros and ones occurring at the PCM bit rate. This would be done, for example, by sampling the sign of the output of an integrate-and-dump filter suitably synchronized with the bit rate and epoch at the output of each receiver. Recall that the PCM information is transmitted in the form of direct carrier biphase (or, more properly, biphase with residual carrier) modulation. The requisite bit synchronizing information can be obtained in any of several ways.

The advantage of making a 1-0 decision at each subaperture receiver is that only binary data need be transmitted to the central point. This fact implies, for example, that the outputs of the various transmission lines need not be provided with any form of automatic gain control. Rather, the outputs can simply be hard limited to provide waveforms which contain all the information transmitted from the neighborhood of the individual subapertures. Use of a hard limiter implies that the zero-one information is encoded as a positive or negative video voltage. Many other options are possible, including modulating a carrier with the binary intelligence and suitably extracting it at the central station. These factors will not be dwelt upon at this time because they are matters of detail. The detail is important in making a final design decision about various alternatives with regard to the factors of cost and reliability, but is not fundamental to showing the basic feasibility of a system of the sort here contemplated.

---

*The detected output of a subarray receiver is referred to as an "analog" signal even though it may be a digital bit stream.

Another possible alternative would be to parallel a number of MOSFET shift registers with the sampling times of the inputs suitably staggered and to combine the outputs through an appropriate logic network. By such means, the variable delay line required could be realized with relatively low speed MOSFET units, each of which would require only a fraction as many stages as for the single shift register device described previously. Even with the discrete component implementation, the variable-delay shift register would be sufficiently inexpensive relative to the total investment in one channel of the multi-aperture receiving system so that cost should not be a significant factor.

Digital implementations for the variable delay devices has been discussed at some length since such an implementation may eventually prove most convenient and economical for the intended application. Such units tend to low cost production when large numbers of identical units are required and are ideally suited for precision delay control.

A detailed design of the digital delay device required would necessitate a more careful analysis of the minimum clock frequency required and would also undoubtedly provide for some additional margin in the total range of delay variability.

The discussion given should, however, be sufficient to establish the feasibility and basic attractiveness of this method of implementing the variable delay device.

Two Methods of Signal Processing. The signals from the individual subapertures must be time delay compensated, phase adjusted, and summed. Time delay compensation and phase summing have been discussed in the previous subsections. This subsection documents two means of performing the signal summation and analyzes the relative performance expected from these two types of signal summation.

Signal Processing Using Bit Decision at the Subaperture. Heretofore it has been supposed that the signals from each subarray have

Signal Processing Using Bit Decision After Summation of the Subaperture Signals. A second option is available for transmitting the information received at the individual subapertures to a common point in digital form. This option consists merely of sampling the analog output of the phase detector from which the noise-corrupted PCM information is obtained and quantizing this information with an analog-to-digital converter. This technique preserves the basic advantage of digital data transmission to the central point, but requires additional equipment. At the central station, the resulting digital data can be converted back into analog form prior to summation. Alternatively, provided that provision is made for synchronizing the words received from the various transmission lines properly, the data received from the various receivers can be added directly in a digital format. After addition it can, if desired, be converted back to an analog signal which would be used to drive a conventional PCM synchronizing and demodulation system. This made of operation supposes, of course, that the time delays are compensated prior to the signal addition, whether the addition be performed in digital or analog form.

Whether the PCM data are transmitted in the form of a single binit for each received bit, from each receiver, or in the form of a PCM word obtained by sampling and quantizing the analog output of the individual receiver, one means of implementing the necessary digital time compensation, addition for signal-to-noise enhancement, and PCM synchronization and demodulation deserves special consideration. Inasmuch as central computer should be employed for control of the entire multi-aperture receiving system, it seems reasonable to consider employing this central control computer to perform the signal processing functions required to demodulate and channelize PCM data. The advantages of digital data handling are so great and the number of inputs sufficiently large that the design of a special purpose computer for performing these functions may well prove to be the preferred implementation for all the signal-processing functions. This possibility warrants considerable additional study.

Signal Processing Analysis

Introduction. For the most part, the performance implications of the multi-aperture receiving system described in the preceding subsection have been readily apparent and have been noted explicitly in the course of the preceding discussion. The one aspect of system performance which requires a somewhat more detailed analysis for an adequate preliminary understanding concerns the possible performance differences between the two essentially distinct ways of handling the data obtained from the phase-locked receivers associated with the individual subapertures. The first way of handling such data consists of transmitting these data to a central station in analog form and simply summing them directly, after correcting the differential time delays, to enhance the output signal-to-noise ratio. The second technique consists of providing a suitable PCM bit synchronizing system at the outlying receiver site and making a bit-for-bit decision as to whether the received PCM bit is a "0" or a "1" at the receiver output. The former case will be refered to as digitizing after correlation and the latter case as digitizing prior to correlation.

While it is obvious that digitizing prior to correlation cannot produce performance as good as that obtained by digitizing after correlation, the amount of degradation which results from digitizing before correlation is not obvious without some analysis. In view of the importance attached to the pre-digitizing, or digitizing prior to correlation, it seems worthwhile to provide a comparative analysis of the two techniques at this point. The discussion accompanying the analysis will also serve to clarify some of the implied details of implementation, which were not specifically indicated in the preceding subsection.

Consider now the modulation output from one of the phase-locked receivers. The modulation format employed is biphase with residual carrier. In addition, a separate timing waveform may be transmitted from the space vehicle to provide a timing reference and bit synchronization signal. In any case, the amount of power required for the

carrier and the timing signal will be so small as to be entirely negligible relative to that remaining in the modulation. Accordingly, to simplify the analysis the residual carrier and the timing signal powers may be ignored and the received signal is assumed to be purely biphase modulated. The modification to account for the presence of the carrier and the timing signal is essentially trivial and amounts to little more than a notational complication.*

As seen at the input to the phase detector in the phase-locked receiver, the received signal and noise may be represented in the form

$$I(t) = \pm A \cos \omega_o t + N_1(t) \cos \omega_o t + N_2(t) \sin \omega_o t \qquad (1)$$

Here the plus or minus sign represents the biphase modulation. The first term is the signal component of the IF input to the phase detector. The last terms are noise terms written in a form which displays the in-phase and quadrature components. The (one-sided) spectral density of the IF noise is taken to be δ_N.** With the in-phase and quadrature representation chosen for the noise, $N_1(t)$ and $N_2(t)$ are statistically independent stationary Gaussian processes with low-pass spectra whose density (one-sided) is $2\delta_N$; the mean values of N_1 and N_2 are zero.

---

*If the maximum phase excursion for the biphase modulation with residual carrier is $\pm\varphi$, the quantity A in Equation 1 below should be replaced by A sin φ. Alternatively, Equation 1 can be used unchanged provided that $A^2/2$ is interpreted as the power in the PCM modulation rather than as the total power of the angle-modulated signal.

**In referring to one-sided spectral densities only positive frequencies are considered. The assumption that the IF noise spectrum is flat is equivalent to the requirement that the IF bandwidth should be sufficiently wide to pass the biphase modulation without significant distortion; this is a reasonable requirement.

The action of the phase detector can be viewed as multiplying I(t) by 2 cos $\omega_0 t$ and taking the low-pass component of the product. In this way the modulation output is obtained as a signal proportional to the quantity V(t) given by

$$V(t) = \pm A + N_1(t) \tag{2}$$

This output is then passed through an appropriate low-pass filter, after which it is either 1) transmitted to the central station in analog form for delay compensation and summation or 2) sampled and a decision made as to whether the received PCM bit was a "0" or a "1."

Digitizing Prior to Correlation. Consider now the case in which the receiver outputs are individually digitized prior to correlation. The optimum output filter will be that which maximizes the output signal-to-noise ratio at the time when the zero-one bit decision is made. With the white noise characteristics here assumed, this optimum (or matched) output filter takes the form of an integrate-and-dump circuit synchronized with the bit period. That is, the modulation output of the phase detector at the beginning of a bit period is gated into an integrating circuit and integrated over the duration of this period. At the end of the period, the voltage across the integrator is read and the integrator is reset to zero. The operation is then repeated for the succeeding bit. A decision as to whether the input bit was a "0" or a "1" is made on the basis of the polarity of the signal read out of the integrate and-dump circuit at the end of the bit period.

If, for example, a bit starts at t = 0 and the output of the integrate-and-dump period is normalized so that it preserves DC levels, its action may be represented by the operator

$$\frac{1}{T}\int_0^T (\quad) dt,$$

where T is the bit period. On applying this operator to the modulation output of the phase detector given by Equation 2, the following is obtained for the output of the integrate-and-dump circuit at the end of the bit period

$$D = \pm A + \frac{1}{T}\int_0^T N_1(t)dt$$

$$= \pm A + X, \tag{3}$$

where, as the integral of a zero-mean Gaussian process, the random variable X has a Gaussian distribution with zero mean.

To calculate the variance of the noise output, X, note that a simple calculation shows that the integrate-and-dump circuit, considered as a frequency filter, has a noise bandwidth, in cps, given by $1/(2T) = F/2$, where F is the bit rate of the received PCM data. Since the frequency characteristic of the integrate-and-dump circuit with the normalizing factor $1/T$ is unity at zero frequency, it follows that output noise power, or equivalently the variance of X, may be calculated by merely multiplying the spectral density of $N_1(t)$ by the noise bandwidth of the integrate-and-dump circuit. Since the spectral density of $N_1(t)$ is $2\delta_N$, the output noise power $\sigma^2(X)$, from the integrate-and-dump circuit is given by

$$\sigma^2(X) = (2\delta_N)\left(\frac{1}{2T}\right) = \frac{\delta_N}{T} = \delta_N F \tag{4}$$

On writing z for the output signal-to-noise ratio from the integrate-and-dump circuit,

$$z = \frac{A^2}{\sigma^2(X)} = \frac{A^2}{\delta_N F} \tag{5}$$

This will be used shortly.

It is also of interest to observe that z can be written in the alternative form

$$z = \frac{A^2/2}{\delta_N F/2} \tag{6}$$

This alternative form for the output signal-to-noise ratio is of some interest as it expresses this ratio in terms of an equivalent signal-to-noise ratio in an appropriate noise bandwidth referred to the intermediate frequency. To see that this is the case, observe that, according to Equation 1, the signal power at IF is given by $A^2/2$ and that the spectral density of the IF noise is δ_N. With the interpretation of $A^2/2$ as the IF signal power, it follows that F/2 is to be interpreted as the equivalent IF noise bandwidth for the combination of the operations of phase detection with respect to the carrier reference and the integrate-and-dump filtering. This interpretation was the basis for the somewhat imprecise statement made in the last subsection to the effect that the output noise bandwidth of the integrate-and-dump circuit in the receiving system was one-half the bit rate, or 500 kilocycles per second for the case discussed there.

The interpretation of Equation 6 given in the last paragraph is sufficiently interesting to warrant further interpretation on physical grounds. The action of the phase detector is to suppress the component of the noise in quadrature with the carrier reference signal and to fold the spectrum of the in-phase component of noise about its center frequency. The result is that the spectrum of the low-pass noise at the phase detector output has, on the one hand, its spectral density halved by suppression of the quadrature component while, on the other hand, the spectral density is doubled by the spectral folding action. It follows from this combination of suppression and folding that the one-sided low-pass noise output spectral density from the phase detector is equal to the spectral density of the original IF noise except for such scale-factor changes as are applied to the signal and the noise alike in

the demodulation process; the detailed derivation given previously takes specific account of these common scale-factor changes. The important point to remember is that the output signal-to-noise ratio obtained from the combination of the phase detector and the integrate-and-dump filter (read at the sampling time) is equivalent to the IF signal-to-noise ratio corresponding to a noise bandwidth equal to one-half of the bit rate.

Since the PCM data is of concern here, the most significant error statistic is not output signal-to-noise but, rather, bit error probability. This is a doubly significant parameter inasmuch as the two distinct methods of processing the received signal data, digitizing prior to correlation and digitizing after correlation, have somewhat different noise characteristics. This fact renders a comparison on the basis of final output signal-to-noise ratios rather meaningless. Fortunately, a direct comparison on the basis of final output bit error probability is feasible. The derivation of suitable expressions for effecting a comparison between the two methods is the major purpose of this subsection.

In the course of the derivations to follow, it will be convenient to use the symbol $P\{\ \}$ for the probability of the event described within the braces.

At a first step, an expression is derived for the probability of a correct bit decision at the output of the integrate-and-dump filter when the digitizing prior to correlation technique is employed. Since the problem is completely symmetrical with respect to 0's and 1's, there is no loss of generality in taking the sign before A in Equation 1 to be positive. With this convention and by using the symmetry of the noise distribution with respect to its zero mean, the probability, p, of a correct decision is:

$$p = P\{D > 0\} = P\{A + X > 0\} = P\left\{\frac{X}{\sigma(X)} > -\frac{A}{\sigma(X)}\right\}$$

$$= P\left\{\frac{X}{\sigma(X)} < \frac{A}{\sigma(X)}\right\} = P\left\{\frac{-X}{\sigma(X)} < \sqrt{z}\right\} \quad . \tag{7}$$

Now $-X/\sigma(X)$ is a normally distributed random variable with mean zero and variance unity. From this observation and Equation 7 it follows that:

$$p = \Phi(\sqrt{z}) , \tag{8}$$

where $\phi(x)$ is the cumulative distribution function for a Gaussian process with zero mean and unit variance, and is given explicitly by

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} \, du . \tag{9}$$

Equation 8 gives the probability of a correct bit decision. It is important to observe that the derivation leading to this equation is dependent only on the signal-to-noise ratio at the time the modulation output of the phase-locked receiver is sampled; that is, it is independent of the particular output filter employed so long as the output signal-to-noise ratio is measured correctly. The integrate-and-dump filter is that filter which will maximize the output signal-to-noise ratio, but not the only output filter which could be used. The derivation of the output signal-to-noise ratio for the integrate-and-dump filter was included to provide a basis for assessing the optimum performance that can be expected from the system. It will be seen that, so long as the same method of making a zero-one bit decision is employed for the digitize-prior-to-correlation system and the digitize-after-correlation system, the relative performance will be independent of the specific output filter chosen; that is, so long as the same output filter and sampling time are employed in the two cases, the comparison will remain valid.

From Equation 8, the probability of bit error, $q = 1-p$, may be calculated with the aid of a table of the cumulative Gaussian distribution. The results of such a calculation, which may be found in various places in the published literature, are shown in Figure E-1 on the following page.

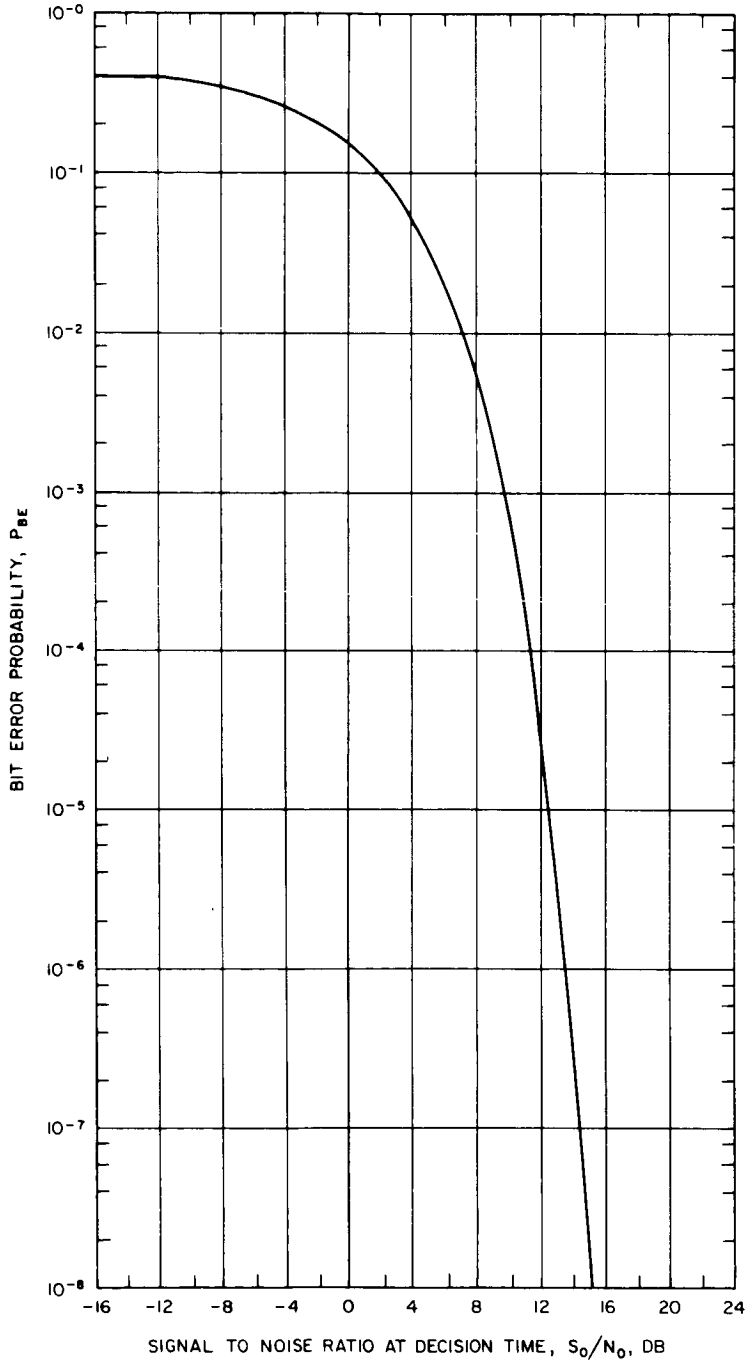


Figure E-1. Bit error probability versus signal to noise power ratio at the decision time.

The probability of making a correct bit decision on the basis of the digitized outputs of all the n receivers will now be made. Here it is assumed that the PCM bit decisions from the various receivers have been appropriately time compensated, so that the data on which the final bit decision are to be based all apply to the same bit. For this purpose, let S_n be the number of correct bit decisions out of the total of n decisions made at the individual receivers. The decisions made at the individual receivers will be statistically independent, since the receiver noise will be independent from one receiver to another and will have common probabilities, p, given by Equation 8. The distribution of S_n will clearly be a binomial distribution corresponding to n independent trials of an event of probability p. The decision logic will consist merely of deciding that the received bit was a "1" if more than half of the n bits received from the individual phase-locked receivers were taken to be 1's, and to be a "0" in the contrary case. In the case where n is even, the decision can be made arbitrarily when precisely half of the received bits were taken to be 1's. Since this will be an event of small probability, the specific choice made for the half-and-half case will not be very important. Since p is the probability of a correct bit decision at the output of an individual phase-locked receiver, the probability that the final zero-or-one decision is correct, when made by the majority vote procedure here indicated, will be given explicitly by

$$P_1 = \sum_{k > \frac{n}{2}} {}_nC_k \, p^k (1-p)^{n-k} \qquad (10)$$

where ${}_nC_k$ is the binomial coefficient (the number of combinations of n things taken k at a time). For present purposes, the more significant quantity is the bit error probability, Q_1, given by

$$Q_1 = 1 - P_1 . \qquad (11)$$

Digitizing After Correlation. To determine the corresponding bit error probability for the system in which the data are digitized after correlation, it is only necessary to observe that the effect of adding the time-compensated outputs of the n transmission lines (with equal gains), filtering, and making a zero-or-one bit decision using the same kind of filter as was supposed previously for the pre-digitizing case is equivalent to increasing the signal-to-noise ratio, z, in Equation 8 by a factor of n. This is the case because the signal components of the outputs from the various receivers are added coherently while the noise components are added incoherently. The fact that this results in an effective increase in signal-to-noise ratio, z, in a factor equal to the number of subapertures employed was used in an essential way in the last subsection and, in fact, constitutes the basis for the entire system design. The probability of bit error for the system which digitizes after correlation is given as

$$P_2 = \Phi(\sqrt{nz}), \tag{12}$$

where z is the output signal-to-noise ratio obtained when a zero-one bit decision is made at the output of an individual phase-locked receiver associated with a single subaperture. The corresponding bit error probability, Q_2, is given by

$$Q_2 = 1 - P_2 . \tag{13}$$

Comparison of the Two Correlation Methods. In principle, the expressions derived for Q_1 and Q_2 provide a direct means of comparing the performance of the two signal-processing techniques for any value of single-receiver output signal-to-noise ratio, z. Such a comparison can be effected by straightforward numerical processes. However, it is desirable to convert the resulting error probability expressions into simpler functional forms which permit a direct performance comparison for the case of primary interest, namely,

that case in which the final output signal-to-noise ratio is such as to provide acceptable bit error rates, but in which the bit error rates which would be obtained at the outputs of the individual phase-locked receivers would be so high as to render the data useless. This is obviously the case of major interest because it represents the performance for which the total receiving system has been designed. Unless the bit error rates at the outputs of the individual phase-locked receivers are so high as to render the data useless, no need exists for the composite receiving system which is the object of this section. The resulting conditions on the single-receiver output signal-to-noise ratio, z, are that z should be small relative to unity, but that nz should be large relative to unity.

Under the conditions described previously, where nz is large relative to unity, we may use the asymptotic form for ϕ. This asymptotic form, valid for large positive x, is given by

$$1 - \Phi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}\,x} \tag{14}$$

From Equations 13 and 14

$$Q_2 = \frac{e^{-(1/2)\,nz}}{\sqrt{2\pi\,nz}} \tag{15}$$

It remains to find a corresponding asymptotic form for Q_1 or, equivalently, P_1. For this purpose observe that the binomial distribution which led to Equation 10 has mean np and variance npq, where q = 1-p. The probability P_1 may now be written as:

$$P_1 = P\left\{S_n > n/2\right\} = P\left\{\frac{S_n - np}{\sqrt{npq}} > -\sqrt{\frac{n}{pq}}\,(p - 1/2)\right\} \tag{16}$$

Now, according to the central limit theorem, or more precisely, the special case applicable to the binomial distribution known as the DeMoivre-Laplace theorem, the distribution of the variable S_n^* given by

$$S_n^* = \frac{S_n - np}{\sqrt{npq}}$$

tends to a Gaussian distribution with mean zero and variance unity; moreover, the convergence is uniform in the sense that $P\{S_n^* \leq x\}$ converges to $\Phi(x)$ uniformly in x.[†] From this fact it follows that

$$P_1 \cong P\left\{Y > -\sqrt{\frac{n}{pq}}\left(p - \frac{1}{2}\right)\right\} = P\left\{-Y < \sqrt{\frac{n}{pq}}\left(p - \frac{1}{2}\right)\right\},$$

where Y is a normally distributed random variable with mean zero and variance unity. By the symmetry of the distribution, -Y has the same distribution, so that the last equation shows that

$$P_1 \cong \Phi\left\{\sqrt{\frac{n}{pq}}\left(p - \frac{1}{2}\right)\right\} \tag{18}$$

From Equation 18,

$$Q_1 \cong 1 - \Phi\left\{\sqrt{\frac{n}{pq}}\left(p - \frac{1}{2}\right)\right\} \tag{19}$$

The result expressed by Equation 19, while correct, requires some explanation. As n tends toward infinity, both Q_1 and the right hand side of Equation 19 tend to zero. The uniformity in x of the convergence asserted by the DeMoivre-Laplace theorem does not, in itself, imply

[†] The DeMoivre-Laplace limit theorem and its generalization, the central limit theorem, may be found treated in some detail in almost any serious text on probability theory. Reference 3 is particularly recommended for the thoroughness and depth of its treatment.

that the ratio of the two sides of Equation 19 tends to unity. This is a mathematical subtlety which requires considerable caution. If the convergence in question is, for the case at hand, such that the ratio of the two sides of Equation 19 tends towards unity, the results that follow will be valid. Since the precise nature of this asymptotic convergence would require a considerably more detailed investigation of rather delicate nature, this question can not be answered at the present time. Rather, the ratio of the two sides in Equation 19 is assumed asymptotic to unity for the present analysis. However, a more detailed investigation of the nature of the convergence should be made before the asymptotic results to be derived are employed as a basis for design. In any event, a comparison can be made by numerical means merely by employing the exact Equation 10.

The bit error probability is small for the cases of interest. In the case of Q_1, only when the argument of Φ in Equation 19 is large relative to unity is of interest. Under these circumstances, the right hand side of Equation 19 can be approximated by use of Equation 14. Before invoking this approximation it will be convenient to develop a more convenient asymptotic expression for the argument of Φ on the right side of Equation 19. For this purpose, Equation 8 is used and recall is made that the signal-to-noise ratio, z, is small relative to unity for the case of interest. For small x, the approximation

$$\Phi(x) = \frac{1}{2} + \frac{x}{\sqrt{2\pi}} \tag{20}$$

may be used in Equation 8, giving

$$p = \frac{1}{2} + \sqrt{\frac{z}{2\pi}} \quad , \quad q = \frac{1}{2} - \sqrt{\frac{z}{2\pi}} \tag{21}$$

Thus, to first order in $\sqrt{z}$,

$$\sqrt{\frac{n}{pq}}\left(p - \frac{1}{2}\right) = \sqrt{\frac{2nz}{\pi}} \tag{22}$$

Recall once again that, for the case of interest, nz is large relative to unity. By combining Equations 14, 19, and 22, the probability of bit error for a system which digitizes prior to correlation is,

$$Q_1 \cong \frac{e^{-\frac{nz}{\pi}}}{2\sqrt{nz}}$$

which can be rewritten in the more useful form

$$Q_1 \cong \frac{e^{-\frac{1}{2} n \left(\frac{2}{\pi} z\right)}}{\sqrt{2\pi n \left(\frac{2}{\pi} z\right)}} \tag{23}$$

Assuming the validity of the approximation made in employing the central limit theorem, it is now a straightforward matter to compare the performance of the two different signal-processing techniques. For this purpose observe that the bit error rate given by Equation 23 has exactly the same functional form as that given by Equation 15 except that the quantity z in Equation 15 is replaced by the quantity $(2/\pi)z$. It follows at once from this observation that the system which digitizes prior to correlation requires a signal-to-noise ratio $\pi/2$ times that required by the system which digitizes following correlation in order to obtain the same final bit error rate. That is, pre-digitizing the data at the output of the phase-locked receiver associated with each individual subaperture requires a signal-to-noise ratio at the receiver output, and hence a received signal strength, which is larger by $10 \log_{10} \pi/2 = 2.0$ db. While every effort should be made in a deep space communications system to obtain maximum communications efficiency, the advantages of pre-digitizing the received data may be sufficient to warrant consideration of this mode of operation in spite of the 2-db performance penalty.

It is of some interest to observe that the performance comparison made above could have been made without the use of the asymptotic form given by Equation 14. It would only have been necessary to introduce the approximateion of Equation 22 into Equation 19 and compare the result with the expression for Q_2 obtainable from Equation 12 by means of Equation 13. Explicit use of the asymptotic expression for $1 - \phi(x)$ was chosen to provide equations which may be conveniently used for numerical calculation of bit error probabilities.

USE OF A COMPUTER

The final point to be made in this subsection is the necessity for computer control of the composite array. The beam steering and signal processing operations required to control properly a large number of electronically-scanned subapertures and to extract the desired data from the received signal are sufficiently complex and numerous that manual control of the system would simply not be feasible. The computer could also perform both time delay compensation and the final zero-or-one decision-making process. Such a computer could also decommutate and smooth the received data and could decode the data if a coding scheme was employed. It goes almost without saying that parity checking and error correction would then be performed in the computer at the same time. These possibilities, while not directly germane to determining the feasibility of the proposed receiving system, are sufficiently attractive to warrant further study.

REFERENCES

1. P. G. Smith, Systems capabilities of arrays of large-aperture antennas," IEEE Transactions on Aerospace and Electronic Systems, vol. AES-2, no. 3, pp. 240-251, May 1966.

2. Analytical study to define an experimental program for the evaluation and optimization of multi-element large aperture arrays, Final Report, RTI Program RU-170, NASA Contract NAS 1-3780, for Langley Research Center, October 1964.

3. B. V. Gnedenko and A. N. Kolmogorov, Limit Distributions for Sums of Independent Random Variables, translated from the Russian and annotated by K. L. Chung, Addison Wesley, Cambridge, Mass., 1954.

APPENDIX F

LASER COMMUNICATION PHOTODETECTION EXPERIMENTS

1. INTRODUCTION

Present designs for laser communication systems are based upon theoretical analyses which assume the detection process to be governed by Poisson statistics. In addition, for deep space applications the designs for laser communication systems are often predicated upon operation with only a few signal photons per pulse or bit. However, no experimental verifications have been made of laser signal detection capability of optical receivers. This appendix outlines such an experimental program.

The experimental objectives of the program are to gather sufficient photodetection data to:

a) statistically characterize the detection process

b) determine the output frequency spectrum

c) verify theoretically derived signal-to-noise ratio

d) verify theoretically derived probability of error

e) evaluate signal detection techniques

2. OPTICAL DETECTION

There are three basic methods of optical detection: direct, heterodyne, and homodyne detection. With direct detection the photodetector output is the demodulated information signal. In a heterodyne detection system the laser carrier is mixed on the photodetector surface with a local oscillator laser. The signal from the photodetector is at the difference frequency between the carrier and local oscillator, and is detected by a conventional radio frequency receiver. Homodyne detection requires that the local oscillator be frequency and phase locked to the carrier. The photodetector output is the demodulated information signal.

In optical communication systems, detection is impaired by: background radiation in the receiver passband; detector shot noise caused by background radiation, the local oscillator, the signal, and detector dark current; and finally thermal noise caused by resistive elements in the receiver. Table F-1 lists the signal-to-noise ratios for the three basic detection methods in terms of photoelectron counts due to the signal, local oscillator, background, and dark current. Terms used in these expressions are defined in the appendix. The relationships between the photoelectron counts and communication system parameters are listed below.

Signal

$$\mu_{S,S} = \frac{d_r^2 \lambda \eta P_L \tau_t \tau_r \tau_c \tau_a}{h c \theta_t^2 R^2}$$

$$\mu_{S,P} = \frac{\mu_{S,S}}{R_P}$$

$$\mu_{S,B} = \frac{\mu_{S,S}}{R_B}$$

Background

$$\mu_{B,S} = \frac{\pi^2}{16} B_i Q_B d_r^2 \eta \theta_r^2 \tau_r \tau_a$$

$$\mu_{B,P} = \frac{\mu_{B,S}}{R_P}$$

$$\mu_{B,B} = \frac{\mu_{B,S}}{R_B}$$

Dark Current

$$\mu_{D,S} = \frac{i_D}{q}$$

$$\mu_{D,P} = \frac{\mu_{D,S}}{R_P}$$

$$\mu_{D,B} = \frac{\mu_{D,S}}{R_B}$$

Local Oscillator

$$\mu_{O,S} = \frac{P_O \eta}{h\nu}$$

$$\mu_{O,P} = \frac{\mu_{O,S}}{R_P}$$

$$\mu_{O,B} = \frac{\mu_{O,S}}{R_B}$$

The transmission capability of analog communication systems is determined by the signal-to-noise ratios of Table F-1; while the capability of time sampled systems is measured by the probability of detection error. To derive the latter quantity it is necessary to know the statistical characteristics of the detection process. If it is assumed that the emissive sources radiate photons into the detector uniformly, then the photoelectron counts of the detector output are given by a Poisson distribution. Table F-2 gives probability of error expressions for various types of modulation and detection methods.

Detection Method	Conditions	Expression
Direct Detection (analog transmission)	$B_o \ll B_i$	$\frac{S}{N} = \frac{\mu_{S,S}^2}{\frac{kTB_o}{q^2G^2R_L} + 2B_o(\mu_{S,S} + \mu_{B,S} + \mu_{D,S})}$
Direct Detection (digital transmission)	$B_o \ll B_i$	$\frac{S}{N} = \frac{\mu_{S,B}^2}{\frac{kTB_o}{q^2G^2R_LR_B^2} + \frac{2B_o}{R_B}(\mu_{S,B} + \mu_{B,B} + \mu_{D,B})}$
Heterodyne Detection (analog transmission)		$\frac{S}{N} = \frac{\mu_{S,S}\mu_{O,S}}{\frac{kTB_o}{q^2G^2R_L} + \mu_{B,S}\mu_{O,S} + B_o(\mu_{S,S} + \mu_{O,S} + \mu_{B,S} + \mu_{D,S})}$
	$\mu_{O,S}$ large	$\frac{S}{N} = \frac{\mu_{S,S}}{B_o + \mu_{B,S}}$
Heterodyne Detection (digital transmission)		$\frac{S}{N} = \frac{\mu_{S,B}\mu_{O,B}}{\frac{kTB_o}{q^2G^2R_LR_B} + \mu_{B,B}\mu_{O,B} + \frac{B_o}{R_B}(\mu_{S,B} + \mu_{O,B} + \mu_{B,B} + \mu_{D,B})}$
	$\mu_{O,B}$ large	$\frac{S}{N} = \frac{\mu_{S,B}}{\frac{B_o}{R_B} + \mu_{B,B}}$

Table F-1. Signal-to-noise ratio expressions.

Detection Method	Conditions	Expression
Homodyne Detection (analog transmission)	$\mu_{O,S}$ large	$\frac{S}{N} = \frac{2\mu_{S,S}}{B_o \mu_{B,S}}$
Homodyne Detection (digital transmission)	$\mu_{O,B}$ large	$\frac{S}{N} = \frac{2\mu_{S,B}}{\frac{B_o}{R_B} + \mu_{B,B}}$

Table F-1. (Continued)

System	Detection Statistics	Expression
PPM (Threshold Detection)	Poisson	$P'_e = \left[1 - \frac{P_{SN}^P}{LP_N^P}\right] + \left[\frac{\left(1-P_N^P\right)^{L-1}}{LP_N^P}\right]\left[P_{SN}^P - P_N^P\right]$ where $P_N^P = \sum_{k=N_T}^{\infty} \frac{\left(\frac{\mu_{N,P}}{L}\right)^k \exp\left\{-\left(\frac{\mu_{N,P}}{L}\right)\right\}}{k!}$ $P_{SN}^P = \sum_{k=N_T^P}^{\infty} \frac{\left(\mu_{S,P} + \frac{\mu_{N,P}}{L}\right)^k \exp\left\{-\left(\mu_{S,P} + \frac{\mu_{N,P}}{L}\right)\right\}}{k!}$ N_T^P = greatest integer value of k_T^P $k_T^P = \frac{\mu_{S,P} + \ln(L-1)}{\ln\left[1 + \frac{L\mu_{S,P}}{\mu_{N,P}}\right]}$

Table F-2. Probability of error expressions.

System	Detection Statistics	Expression
PPM (Majority Detection)	Poisson	$P_e' = 1 - \left\{ \sum_{n=o}^{\infty} \sum_{m=o}^{\infty} \frac{\left(\mu_{S,P} + \frac{\mu_{N,P}}{L}\right)^{n+m} \left(\frac{\mu_{N,P}}{L}\right)^{m} \exp\left\{ -\left(\mu_{S,P} + \frac{2\mu_{N,P}}{L}\right)\right\}}{(n+m)!\, m!} \right\}$ $+ \frac{L-1}{L} \left\{ \sum_{m=o}^{\infty} \frac{\left(\mu_{S,P} + \frac{\mu_{N,P}}{L}\right)^{m} \left(\frac{\mu_{N,P}}{L}\right)^{m} \exp\left\{ -\left(\mu_{S,P} + \frac{2\mu_{N,P}}{L}\right)\right\}}{m!\, m!} \right\}$
PCM/IM (Threshold Detection)	Poisson	$P_e = 1/2\,(1 + P_N^B - P_{SN}^B)$ where $P_N^B = \sum_{k=N_T^B}^{\infty} \frac{\left(\frac{\mu_{N,B}}{L}\right)^{k} \exp\left\{ -\left(\frac{\mu_{N,B}}{L}\right)\right\}}{k!}$ $P_{SN}^B = \sum_{k=N_T^B}^{\infty} \frac{\left(\mu_{S,B} + \mu_{N,B}\right)^{k} \exp\left\{ -\left(\mu_{S,B} + \mu_{N,B}\right)\right\}}{k!}$ where N_T = greatest interger value of k_T^B $k_T^B = \frac{\mu_{S,B}}{\ln\left[1 + \frac{\mu_{S,B}}{\mu_{N,B}}\right]}$

Table F-2. (Continued)

System	Detection Statistics	Expression
PCM/PL	Poisson	$P_e = 1 - \exp\left\{\left(\mu_{S,B} + \mu_{N,B}\right)\right\} \sum_{n=o}^{\infty} \sum_{m=o}^{\infty} \frac{\left(\mu_{S,B} + \frac{\mu_{N,B}}{2}\right)^{n+m} \left(\frac{\mu_{N,B}}{2}\right)^{m}}{(n+m)!\, m!}$ $+ 1/2 \exp\left\{-\left(\mu_{S,B} + \mu_{N,B}\right)\right\} \sum_{m=o}^{\infty} \frac{\left(\mu_{S,B} + \frac{\mu_{N,B}}{2}\right)^{m} \left(\frac{\mu_{N,B}}{2}\right)^{m}}{m!\, m!}$
PCM/PL	Gaussian Approximation	$P_e = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\frac{\mu_{S,B}}{\sqrt{\mu_{S,B} + \mu_{N,B}}}} \exp\left\{-\left(\frac{W^2}{2}\right)\right\} dW$
PCM/FM (Heterodyne Detection)	Gaussian Approximation	$P_e = 1/2 \exp\left\{-\frac{\mu_{S,B}^2}{\mu_{S,B} + \mu_{N,B}}\right\}$
PCM/PM (Homodyne Detection)	Gaussian Approximation	$P_e = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\frac{\mu_{S,B}}{\sqrt{\mu_{S,B} + \mu_{N,B}}}} \exp\left\{-\left(\frac{W^2}{2}\right)\right\} dW$

Table F-2. (Continued)

3. DESCRIPTION OF EXPERIMENTS

Figure F-1 contains a block diagram of the experimental apparatus for direct, heterodyne, and homodyne photodetection experiments. A variety of data sources feed a signal coder which in turn drives an electro-optic intensity modulator. With these data sources the following types of signal formats may be considered.

No modulation
Sine wave intensity modulation
Pulse intensity modulation
Square wave intensity modulation
Digital intensity modulation

These modulation formats include all practical forms of intensity modulation such as analog intensity modulation (IM), pulse position modulation PPM, and pulse code intensity modulation PCM/IM. Results with these signal formats also indirectly cover the detection of frequency, polarization, and phase modulated carriers, since at present the only feasible means of detecting such signal modulation is by indirect intensity photodetection. For example, PCM polarization modulation is detected by separating the polarization components by a prism, and directing them towards photodetectors.

In the experimental systems provisions are made to examine the output of the photodetector directly, and to decode the received signal. For the latter case, the transmitted and received signals are compared to determine the probability of detection error.

Measurements of signal and noise power will be performed by monitoring the receiver output for the controlled operating conditions listed in the Table F-3.

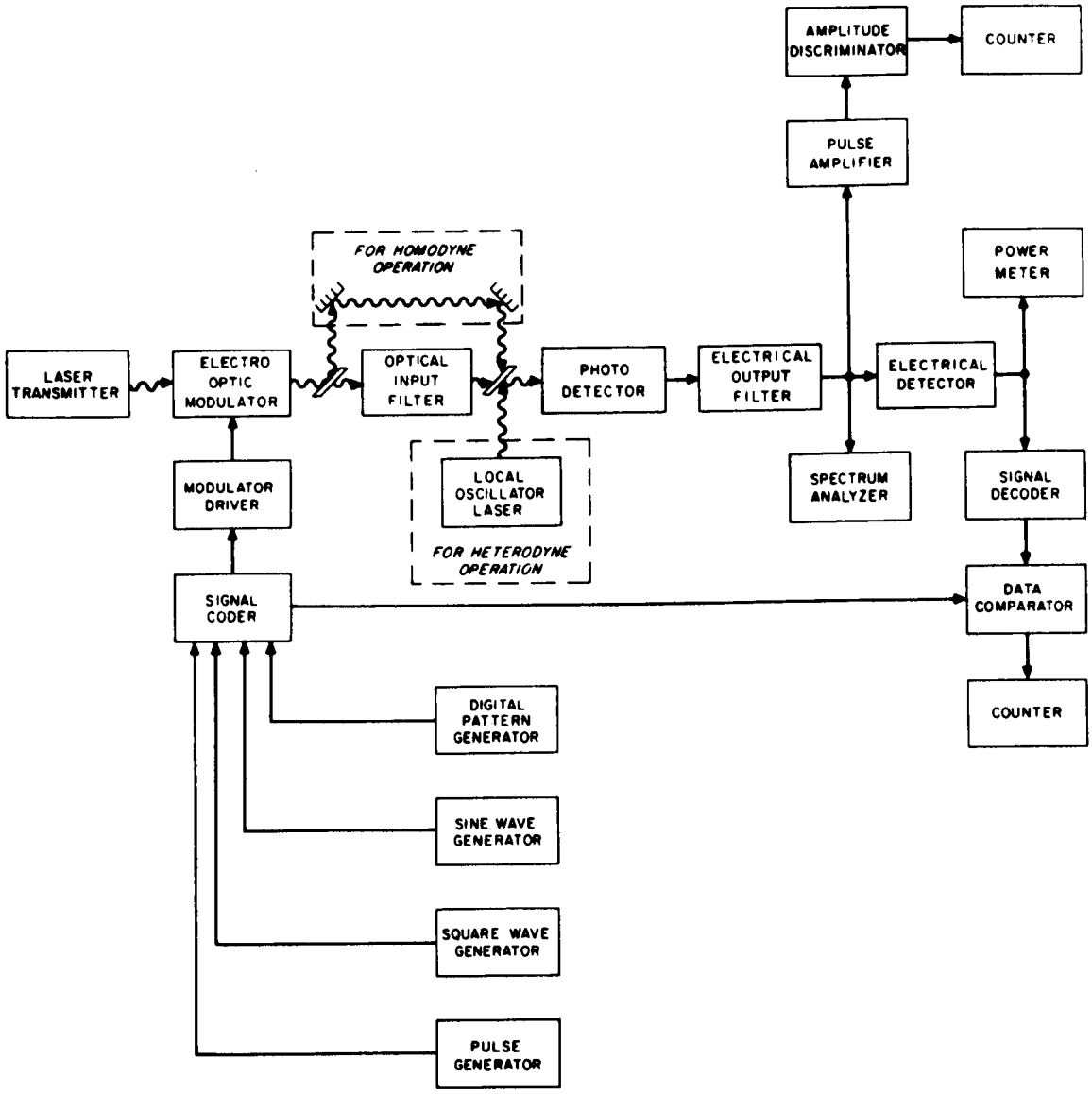


Figure F-1. Experimental optical detector system.

Signal		X X X X X X X X	X X X X X X X X
Local Oscillator			X X X X X X X X
Background	X X X	X X X X	X X X X
Dark Current	X X X	X X X X	X X X X
Thermal	X X X	X X X X	X X X X
Test No.	1 2 3 4 5 6 7	8 9 10 11 12 13 14 15	16 17 18 19 20 21 22 23
Test Type	Photodetector	Direct Detection	Heterodyne-Homodyne Detection

Table F-3. Experimental conditions.

The X's indicate the presence of a signal, local oscillator, or certain type of noise. For example, in Test No. 1, the output of a photodetector is examined without any laser input, with the dark current and thermal noise suppressed by cooling, and with the detector shielded from background radiation. The results of this test will represent the best that can be achieved in "noise free" operation with the experimental apparatus, and further provide a reference for the other tests. In the subsequent tests the signal, local oscillator and the noise effects will be considered singly and in all practical combinations.

The most difficult part of the experiments will be to determine the photoelectron counts due to the arrival of signal and background photons. With a photomultiplier detector each photon which dislodges an electron from the photoemissive surface will produce an output current pulse of magnitude

$$i = \frac{Gq}{\tau}$$

where

G = photodetector multiplication gain

q = electronic charge

τ = detector resolving time

For a typical high quality photomultiplier, $G = 10^6$ and $\tau = 10^{-8}$ seconds so that current pulses of 16×10^{-6} amps will be produced. The photomultiplier acts as an almost ideal current source with a shunt capacitance of approximately 10 pico-farads feeding a load resistor of about 75 ohms as shown below.

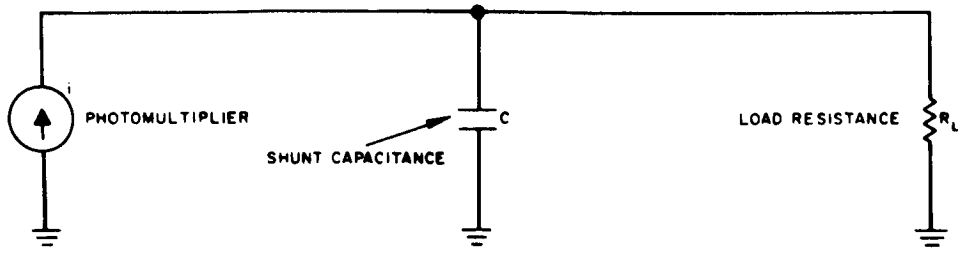


Figure F-2. Photomultiplier output equivalent circuit.

The bandwidth of the output circuit is

$$B_o = \frac{1}{2\pi R_L C} \simeq 2 \times 10^8 \text{ H}_z$$

Thus, since $B_o \simeq \frac{1}{2\tau}$ the detection system should pass the photomultiplier current pulses with minimal pulse shape degradation. The output signal-to-noise ratio (SNR) is from Table F-1.

$$\frac{S}{N} = \frac{\mu_{S^2, B}}{\frac{kTB_o}{q^2 G^2 R_L R_B{}^2} + \frac{2B_o}{R_B}\left(\mu_{S, B} + \mu_{B, B} + \mu_{D, B}\right)}$$

If the signal photon arrival rate is set so that the current pulses appear every τ seconds, the SNR becomes for $B_o = \frac{1}{2\tau}$ is

$$\frac{S}{N} = \frac{\mu_{S,B}^2}{\frac{kT\tau}{2q^2G^2R} + (\mu_{S,B} + \mu_{B,B} + \mu_{D,B})}$$

At a temperature of 300° K the thermal noise term is negligible because of the multiplication gain. Assuming the background and dark current photoelectrons counts to be small the SNR is

$$\frac{S}{N} = \mu_{S,B}$$

But, it has been assumed that only a single photon will be released in the detector resolving time τ so that the SNR is unity.

A similar analysis for heterodyne detection yields a SNR of

$$\frac{S}{N} = \frac{2\mu_{O,B}}{(1 + \mu_{O,B})}$$

for a signal photon arrival rate of $\frac{1}{\tau}$. If the local oscillator photon rate is also set at $\frac{1}{\tau}$ so that a signal-local oscillator photon pair match is made, the SNR becomes unity as in the case of direct detection.

For homodyne detection the SNR for single current pulses is

$$\frac{S}{N} = \frac{4\mu_{O,B}}{(1 + \mu_{O,B})}$$

If the local oscillator rate matches the signal rate the SNR is two.

Because of the low signal-to-noise ratio for single photon detection, it will be necessary to modulate the signal photon rate and apply correlation techniques between the single event detections. The difficulties inherent with monitoring the signal and local oscillator rates, and eliminating dark current, background radiation, and thermal noise may make it necessary to infer information of single photon arrivals from sequential tests in which N and N+1 photons are detected. The signal-to-noise ratio in this experiment approaches N.

GLOSSARY OF TERMS

IM Intensity modulation

FM Frequency modulation

PM Phase modulation

PL Polarization modulation

PPM Pulse position modulation

PCM Pulse code modulation

$\mu_{S,S}$ Average number of signal photoelectrons per second

$\mu_{S,P}$ Average number of signal photoelectrons per sample period

$\mu_{S,B}$ Average number of signal photoelectrons per bit

$\mu_{B,S}$ Average number of background photoelectrons per second

$\mu_{B,P}$ Average number of background photoelectrons per sample period

$\mu_{B,B}$ Average number of background photoelectrons per bit

$\mu_{D,S}$ Average number of dark current photoelectrons per second

$\mu_{D,P}$ Average number of dark current photoelectrons per sample period

$\mu_{D,B}$ Average number of dark current photoelectrons per bit

$\mu_{O,S}$ Average number of local oscillator photoelectrons per second

$\mu_{O,P}$ Average number of local oscillator photoelectrons per sample period

$\mu_{O,B}$ Average number of local oscillator photoelectrons per bit

$\mu_{N,P} = \mu_{B,P} + \mu_{D,P}$

$\mu_{N,B} = \mu_{B,B} + \mu_{D,B}$

i Photodetector current

R_P Information rate (samples per second)

R_B Information rate (bits per second)

B_i Predetection filter bandwidth

B_o Post detection filter bandwidth

τ	Detector resolving time
L	Number of levels of PPM
P_e	Bit error probability
P'_e	Sample error probability
τ_t	Transmitter transmittance
τ_r	Receiver transmittance
τ_a	Atmospheric transmittance
τ_c	Scintillation transmittance
η	Photodetector quantum efficiency
λ	Wavelength
ν	Frequency
θ_t	Transmitter beamwidth
θ_r	Receiver field-of-view
d_r	Receiver aperture diameter
R	Range
P_L	Laser transmitter power
P_O	Local oscillator power incident on detector
$\frac{S}{N}$	Power signal-to-noise ratio
q	Electronic charge
h	Planck's constant
k	Boltzmann's constant
R_L	Effective thermal load resistance
C	Shunt capacitance
G	Photodetector multiplication gain
c	Velocity of light
T	Detector effective noise temperature
Q_B	Background photon spectral radiance
i_D	Photomultiplier dark current